Social Surprises and Feedback Shape Momentary Mood and Anxiety in Humans

Elena Bagdades[1,2†], Marjan Biria[2,6†], Charlotte Burman[2], Raphaelle Delpech[2,6], Quentin J. M. Huys[2,7,11], Madeleine Moses-Payne[2], Jessica Norman[1,2], Diego A. Pizzagalli[8,9], Lucienne Spencer[10], Naomi Tromans[2,3], Ilina Singh[10], Eleanor Leigh[3,5], Georgina Krebs[1], and Argyris Stringaris[1,2,4]

[1] Department of Clinical, Educational & Health Psychology, Division of Psychology and Language Sciences, University College London

[2] Division of Psychiatry, University College London

[3] Department of Experimental Psychology, University of Oxford, Oxford, UK

[4] Department of Psychiatry, National and Kapodistrian University of Athens, Greece

[5] Oxford Health NHS Foundation Trust, Oxford, UK

[6] Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK

[7] Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK

[8] McLean Hospital, Harvard Medical School, Belmont, MA, USA

[9] Department of Psychiatry, Harvard Medical School, Boston, MA, USA

[10] Department of Psychiatry, University of Oxford, Oxford, UK

[11] Department of Imaging Neuroscience, Queen Square Institute of Neurology, University College London, London, UK

† *These authors contributed equally to this work.*

Corresponding author: Elena Bagdades, elena.bagdades.21@ucl.ac.uk

Abstract

Social interactions strongly influence human affect, but the mechanisms linking them to momentary mood and anxiety remain unclear. Building on work showing that reward prediction errors (PEs) shape mood, we extend this framework to the social domain. In a preregistered study, 185 participants (ages 14–45) completed a novel task involving serial social interactions. We compared computational models to test how social feedback, expectations, and PEs related to momentary affect. Results showed that social feedback and PEs best explained affective fluctuations. Social PEs predicted anxiety more strongly than social feedback; the opposite pattern was observed for mood. Further analyses indicated that higher social anxiety symptoms heightened affective responses to social feedback, whereas depressive symptoms increased sensitivity to negative PEs for mood and to social feedback for anxiety. These findings provide a computational account of social affective dynamics across development and highlight novel opportunities for potential risk markers and treatment targets.

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 affect remain surprisingly 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 and anxiety. We reverse-translate elements of clinical theory and observation by introducing a novel experimental and computational framework. Through this framework, we test the role of surprises in social interactions on momentary affect in adolescents and adults, including those 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 the chances of species 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 forms of psychopathology 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, in both health and psychopathology.

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. Specifically, research has highlighted a central role for surprises—or reward prediction errors (RPEs), the discrepancy between expected and actual reward outcomes—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 (Rutledge et al., 2014; Shepperd & Mcnulty, 2002). 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. The literature is mixed, however, on what the effect of RPEs are, over and above that of other appraisals, such as reward magnitude or reward expectation (Forbes & Bennett, 2023; Keren et al., 2021; Rutledge et al., 2014). Computational modelling provides a powerful tool for capturing how such appraisals shape momentary affect, 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; a phenomenon known as the recency bias (Keren et al., 2021). 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 affect has focused on the happy–unhappy spectrum, while other types of affect are largely understudied. Specifically, while the processes involved in pathological anxiety, such as in SAD have been extensively studied, the computations involved in momentary anxiety remain poorly understood. Clinical theory offers valuable insight here. The notion of belief disconfirmation, closely related to prediction error, is central to both exposure therapy, which reduces anxiety through repeated confrontation with feared stimuli (Craske et al., 2014; Salkovskis et al., 2007), and cognitive therapy, which targets and restructures maladaptive beliefs (Clark & Wells, 1995). For example, cognitive therapy for SAD uses behavioural experiments—planned activities designed to test feared predictions—to help individuals compare their expectations with the actual outcomes of social interactions, allowing them to update negative beliefs (Clark, 1999; Leigh & Clark, 2023). This therapeutic principle can be ‘reverse-translated’ to guide experimental investigations into the mechanisms of momentary anxiety, particularly within social contexts. Furthermore, computational modelling of momentary affect provides a powerful framework for examining whether these mechanisms are influenced by internalizing symptoms—such as depression and social anxiety—and for determining the

specificity of such effects; that is, whether they are uniquely associated with one symptom profile or shared across both.

In this study, we addressed key gaps in the literature by examining the computations underlying momentary mood and anxiety in the context of social interactions. Specifically, we designed a novel experimental setup emulating socially demanding situations, in which expectations and outcomes are manipulated, we extended computational models of mood into the social domain and incorporated the measurement of anxious affect in addition to mood. Finally, we investigated how these processes vary across symptom dimensions. To maximise the generalisability of our findings, we recruited a broad age range, including adults aged 18-45 years through the local community and Prolific, as well as adolescents aged 14-18 years through schools. Adolescents were specifically included given the heightened salience of social rewards and punishments in this age group (Foulkes & Blakemore, 2016) and the rising prevalence of mental health problems related to interpersonal difficulties (Collishaw, 2015; Lamblin et al., 2017). This task was co-designed with a youth advisory group to ensure relevance, 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 momentary mood and negatively with momentary anxiety. 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 affect is shaped by immediate social feedback, as well as by the history of positive and negative social PEs. Notably, the accumulated history of social PEs was more predictive of momentary anxiety than social feedback, whereas the opposite pattern emerged for momentary mood. 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 affect. In particular, they showed a greater elation of mood and a larger drop in anxiety in response to positive feedback, and a greater decrease in mood and increase in anxiety in response to negative feedback. Further,

individuals with higher depressive symptoms were characterised by a bigger effect of negative social PEs on their mood and a bigger effect of social feedback on their anxiety. This could suggest that individuals at risk for social anxiety and depression are more susceptible to social appraisals when reporting their momentary affect. Collectively, these findings offer novel insight into how affective dynamics may become dysregulated in psychopathology and point to potential early markers of risk.

Results

Participants

We recruited 722 participants aged 18–25 through Prolific (www.prolific.com) across 21 pilot studies. Data from these pilots were used to refine the experimental task, inform the design of computational models based on preliminary mixed-effects model analyses, and conduct power calculations to estimate the required sample size. A summary of the results from the pilot studies can be found in our preregistration (https://osf.io/73zsg/).

The main study reported here involved a separate sample of 185 participants aged 14–45 years, recruited from three sources: Prolific (n = 97), the University College London community (n = 43), and secondary schools (n = 44) in the UK.

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-SR; 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)).

Participants then completed the Social Prediction, Evaluation and Affect tasK (SPEAK), an online task simulating socially evaluative interactions (see Figure 1). They were told that four different "virtual players" would evaluate how well they came across after describing emotionally neutral 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 an image for 15 seconds and received feedback from the virtual player on a scale from 0 to 100. Immediately after feedback, participants reported their momentary mood and anxiety.

The virtual players varied in their level of harshness, and we manipulated feedback such that, on some trials, it was shifted above or below each virtual player's typical rating behaviour (operationalised as their histogram mean; see Figure 1). This was done with the aim of generating a range of social prediction errors (PEs), including negative (PE^{-}), positive (PE^{+}), strong positive (PE^{++}), and neutral (PE^{N}). In other words, feedback sometimes matched the histogram mean (PE^{N}), and sometimes deviated from it in the positive or negative direction to elicit different magnitudes of PEs. Note that these PE categories reflect our intended experimental manipulation, rather than participants' actual experienced PEs (see Computational Modelling section in Methods). Trial order, feedback, and stimuli were randomised.

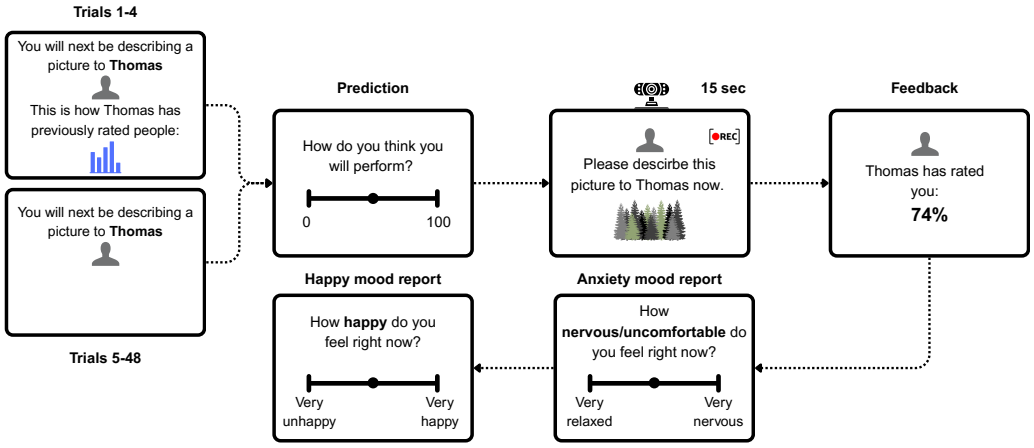

Figure 1. **Structure of a trial in the SPEAK task.** Participants interacted with four different virtual players throughout the task. During the first four trials, each virtual player was introduced alongside a histogram illustrating their rating tendencies, expressed as percentages ranging from 0% to 100%, designed to induce expectations about how critical each virtual 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 mood and anxiety.

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

Out of the eight computational models compared, the best-fitting one for both momentary mood and anxiety included an effect of social feedback and asymmetric effects for positive and negative PEs (see Figure 2). 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, PEs were computed as the difference between the outcome (O; i.e., the amount of social feedback received) and the participant-reported expectation on that trial. Self-reported mood and anxiety at trial t for individual i was modelled as:

$$Mood_{it} = \beta_{0i} + \beta_{1i} \cdot \sum_{j=1}^{t} \gamma^{t-j} \cdot PE_{pos,j} + \beta_{2i} \cdot \sum_{j=1}^{t} \gamma^{t-j} \cdot PE_{neg,j} + \beta_{3i} \cdot O_t \qquad (1)$$

$$Anxiety_{it} = \beta_{0i} + \beta_{1i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot PE_{pos,j} + \beta_{2i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot PE_{neg,j} + \beta_{3i} \cdot O_t \qquad (2)$$

where β_{0i} represents the individual-specific baseline of mood or anxiety, and β_{1i} and β_{2i} capture the sensitivity of individual i to the weighted history of positive and negative PEs, respectively. The parameter γ_i^{t-j} is a forgetting factor that determines the influence of prior trials on affect. O_t represents the social feedback on trial t, and β_{3i} captures its immediate effect on momentary affect for participant i.

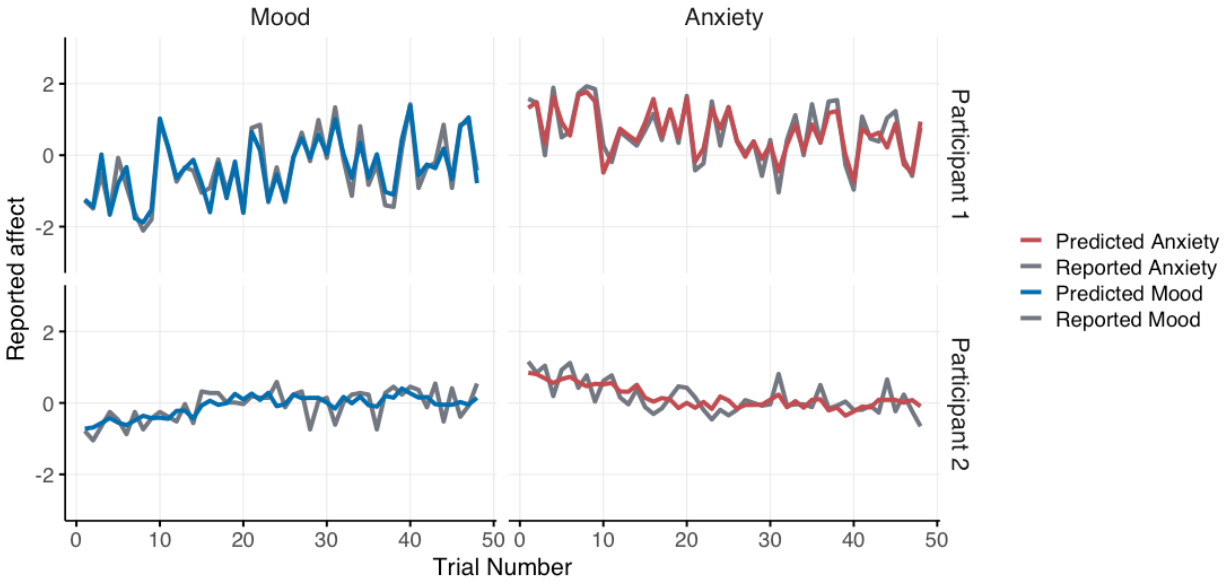

Figure 2. **Reported and predicted momentary mood and anxiety across trials for two example participants.** Each panel shows z-scored values across trials, with mood plotted in the left column and anxiety in the right column. Rows represent individual participants. Reported responses are shown in grey,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

Group-Level Parameter Estimates

Positive Prediction Error Effects (β_1). The effect of positive PEs varied across mood and anxiety models. For momentary mood, the effect was negligible (mean β_1 = 0.05, 95% HDI [-0.02, 0.13]). However, for momentary anxiety, there was a substantial negative effect of positive PEs (mean β_1 = -0.18, 95% HDI [-0.25, -0.10]), indicating that the accumulation of positive surprises was linked to reduced anxiety.

Negative Prediction Error Effects (β_2). Negative PEs did not show an association with momentary mood (mean β_2 = -0.02, 95% HDI [-0.06, 0.03]). For momentary anxiety, however, there was a positive effect (mean β_2 = 0.13, 95% HDI [0.07, 0.18]), suggesting that the accumulation of worse-than-expected outcomes were associated with increased anxiety.

Social Feedback Effects (β_3). Social feedback had a substantial effect on both mood and anxiety. For momentary mood, social feedback had a substantial positive effect (mean β_3 = 0.22, 95% HDI [0.19, 0.26]), indicating that more positive social feedback was associated with improved mood. In contrast, for momentary anxiety, social feedback exerted a negative effect (mean β_3 = -0.09, 95% HDI [-0.11, -0.07]), suggesting that more positive social feedback was associated with reduced anxiety.

Note on parameter comparability. Although all predictors were standardised to facilitate interpretation, the PE coefficients (β_1, β_2) and social feedback coefficient (β_3) are not directly comparable. Social feedback reflects trial-level effects, whereas PEs represent the weighted accumulation of prediction errors across trials (see Supplement for standardisation details). Direct comparison of standardised coefficients would therefore conflate temporally distinct processes. To assess the relative contribution of these predictors to mood and anxiety, we instead compared nested models with and without each predictor (see Supplement).

Forgetting Factor (γ). The forgetting factor γ determines how much influence past prediction errors (PEs) have on current affect, with values bounded between 0 and 1. A value of $\gamma = 0$ implies full recency, where only the most recent trial influences affect, while $\gamma = 1$ implies equal weighting of all past trials, regardless of their temporal distance.

The group-level estimate of γ was 0.93 for mood (95% HDI [0.91, 0.95]) and 0.93 for anxiety (95% HDI [0.92, 0.95]), indicating that past trials have a long-lasting impact on current affect.

These results partially support our hypothesis that social surprises, operationalised 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 momentary mood at the group level. In contrast, PEs had a substantial effect on momentary anxiety that exceeded the effect of social feedback, as confirmed by nested model comparisons of the reduced models (see supplement).

Moderation of effects by social anxiety symptoms

To assess whether social anxiety symptoms (SA), as measured by the LSAS (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 mood and anxiety. For both types of affect, the interaction term for social feedback, $\beta_{3 \cdot SA}$, showed the most robust effect. For anxiety, this interaction was negative (mean = -0.03, 95% HDI [-0.05, -0.02]; see Figure 3 B), indicating that social feedback exerted a stronger influence on anxiety in individuals with elevated SA: more positive feedback was associated with a larger reduction in anxiety, while more negative feedback was associated with a larger increase. In contrast, for mood, the interaction was positive (mean = 0.03, 95% HDI [0, 0.06]; see Figure 3 A), suggesting that the same pattern held for mood, with SA amplifying both the mood-enhancing effects of positive feedback and the mood-lowering effects of negative feedback.

Other interaction terms between SA and the intercept ($\beta_{0 \cdot SA}$), positive PEs ($\beta_{1 \cdot SA}$), negative PEs ($\beta_{2 \cdot SA}$), and the forgetting factor ($\beta_{\gamma \cdot SA}$) 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 [-0.47, 0.91]) and mood (mean = -0.16, 95% HDI [-0.81, 0.54]) were also uncertain.

To complement the model-based findings, we computed the posterior distribution of Pearson correlations between each participant-level parameter estimate and SA scores. The parameter for social feedback (β_3) showed a strong negative correlation with SA in predicting anxiety (mean r = -0.27, 95% CI [-0.33, -0.21], Pr(r < 0) = 100%; see Figure 4 B), and a moderate positive correlation in predicting mood (mean r = 0.13, 95% CI [0.08, 0.17], Pr(r > 0) = 100%; see Figure 4 A). These results provide further support for a moderation of the influence of social feedback on momentary 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 ($\beta_{0 \cdot DEP}$), positive PEs ($\beta_{1 \cdot DEP}$), negative PEs ($\beta_{2 \cdot DEP}$), social feedback ($\beta_{3 \cdot DEP}$) and the forgetting factor ($\beta_{\gamma \cdot DEP}$) had credible intervals that overlapped zero, suggesting inconclusive or negligible moderation effects. Similarly, the main effects of DEP on momentary anxiety (mean = 0.16, 95% HDI [-0.56, 0.84]) and mood (mean = -0.19, 95% HDI [-0.89, 0.49]) were also uncertain.

We also computed the posterior distribution of Pearson correlations between each participant-level parameter estimate and DEP scores. The parameter for social feedback (β_3) showed a negative correlation with DEP in predicting anxiety (mean r = -0.19, 95% CI [-0.26, -0.12], Pr(r < 0) = 100 %; see Figure 4 D), indicating that individuals with higher DEP experienced a stronger anxiety response to social feedback (i.e., larger reductions following more positive feedback and larger increases following more negative feedback). Similarly, the parameter for negative PEs (β_2) showed a negative correlation with DEP in predicting mood (mean r = -0.18, 95% CI [-0.34, 0], Pr(r < 0) = 97.9%; see Figure 4 C). This indicates that individuals with higher DEP tended to exhibit a stronger negative effect of negative PEs on their mood, although the credible interval includes 0. Correlations for all other parameters were weaker and had posterior intervals that crossed zero.

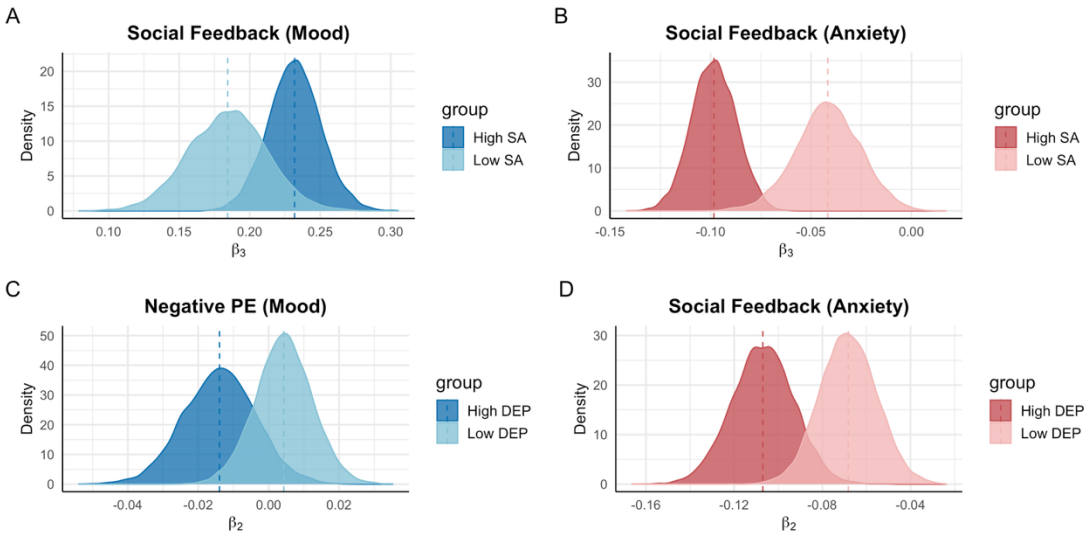

Figure 3. **Posterior distributions by symptom group for each analysis**. (A) Social feedback predicting momentary mood by SA group, (B) Social feedback predicting momentary anxiety by SA group, (C) Negative PE predicting momentary mood by DEP group, and (D) Social feedback predicting momentary anxiety by DEP group. Dashed lines indicate the mean of the posterior distribution for each group. Participants were categorised into high versus low symptom groups based on clinical thresholds, although symptom measures were treated as continuous variables in all models.

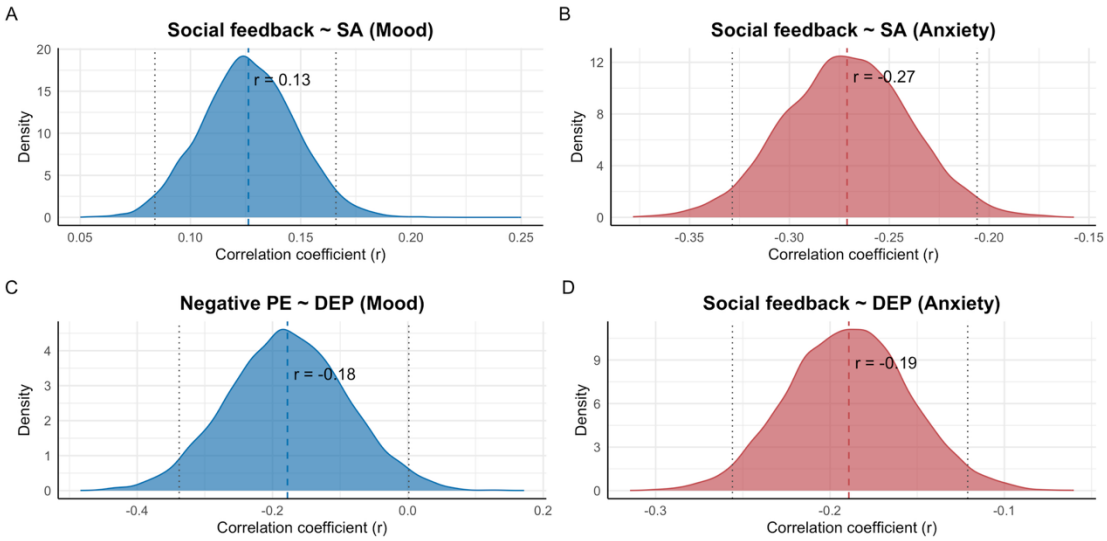

Figure 4. **Posterior distributions of Pearson correlations between participant-level parameter estimates and symptom scores.** (A) Social feedback and SA scores for momentary mood, (B) Social feedback and SA scores for momentary anxiety, (C) Negative PE and DEP scores for momentary mood, and (D) Social feedback and DEP scores for momentary anxiety. Dashed lines indicate the mean Pearson correlation coefficient (r), and dotted lines represent the 95% highest density interval (HDI).

Discussion

In this study, we set out to examine how momentary mood and anxiety are influenced during social interactions. Building on clinical theory and computational models of reward-based mood dynamics, we developed a novel experimental paradigm to elicit and quantify social PEs in

real time. We found that the best-fitting model for both momentary mood and anxiety included social surprises and social feedback, suggesting that both types of social appraisals contribute to shaping affect. This finding provides at least partial support for the central hypothesis that social surprises play a meaningful role in momentary affect. However, our results reveal an asymmetry: momentary mood was more strongly and selectively modulated by immediate social feedback, whereas anxiety was shaped by both social feedback and the cumulative impact of social prediction errors (see Supplement), suggesting distinct computational mechanisms for happy and anxious affect. This pattern may suggest that, in the context of social interactions, mood is more influenced by the valence of social evaluations, while anxiety also reflects sensitivity to expectation violations. These findings underscore the value of a computational approach in uncovering how distinct affective processes are differentially influenced by social experience. Furthermore, we found that individuals with elevated social anxiety symptoms were more sensitive to social feedback in both mood and anxiety, and that depressive symptoms moderated the influence of negative PEs on mood and of social feedback on anxiety. Together, these findings advance our understanding of the computational underpinnings of momentary affect in social contexts and may help identify potential mechanisms by which mood and anxiety dynamics go awry in psychopathology.

Our findings build on and extend influential computational models of mood that conceptualize affect as a function of recently experienced deviations from our expectations, or prediction errors (Rutledge et al., 2014). One theoretical account proposes that mood acts as a momentum signal, integrating over sequences of PEs to track environmental trends and support adaptive behaviour in dynamic contexts (Eldar et al., 2016). Unlike the model proposed by Rutledge et al. (2014), which includes terms for expected reward and a single undifferentiated PE signal, our model incorporates both outcome magnitude and asymmetric positive and negative PEs. Notably, our modelling approach aligns closely with that of Forbes & Bennett (2023), who also found that the best-fitting model of affect included asymmetric PEs and reward outcome magnitude. In their study, as in ours, the inclusion of an outcome term substantially reduced the effect of PEs on mood. However, a key contribution of our work is that we extended these models to momentary anxiety.

While PEs had a minimal effect on mood after accounting for outcomes, they remained substantial predictors of momentary anxiety, suggesting that anxiety may be particularly sensitive to violations of social expectations, above and beyond the valence of the feedback itself. In fact, we found that PE terms contribute more to explaining momentary anxiety than social feedback does, as evidenced by comparing our full model to reduced models (i.e., models with individual predictor terms systematically removed) using fit metrics reported in the Supplement. In this way, we also provide experimental support for the notion that belief disconfirmation during social interactions modulates anxiety, as proposed in clinical theories of SAD (Clark & Wells, 1995). In contrast, mood was more strongly driven by current social feedback than by PEs, and to a greater extent than anxiety. This divergence may reflect distinct functional roles of mood and anxiety, at least in the social domain. Mood may signal how rewarding the environment is, in this case, how well a social interaction is going, while anxiety may serve as a vigilance signal, with its sensitivity to expectation violations tracking uncertainty and ensuring that anticipated social threats are continually monitored and updated. This interpretation is consistent with evolutionary theories that assign distinct adaptive functions to different affective states. According to these accounts, mood acts as a broad regulatory signal that reflects the overall favourability of the environment, guiding approach or withdrawal behaviour, whereas anxiety is specialised for detecting and responding to uncertainty or potential threat, thereby promoting caution and increased vigilance (Marks & Nesse, 1994; Nesse, 1990).

Furthermore, the observed asymmetry may also reflect differences in the temporal dynamics of momentary mood and anxiety. Our modelling results revealed that anxiety showed significant integration of social prediction errors over time (with a forgetting factor close to 1), whereas mood, despite having a similarly high forgetting factor, was primarily driven by immediate trial-level social feedback. This pattern aligns with research on affective chronometry, which emphasizes that distinct emotional states unfold along different temporal profiles, including differences in rise time, duration, and recovery (Davidson, 1998, 2015; Fan et al., 2019). Supporting this, a recent ecological momentary assessment (EMA) study in university students by Villano et al. (2020) found that prediction errors about exam grades had longer-lasting effects on negative affect, including anxiety, than on positive affect such as happiness, further suggesting that anxious states may evolve more slowly and persist longer than momentary mood. Future research should directly investigate temporal dynamics of real-world momentary mood and

anxiety, ideally by combining computational modelling with high-temporal-resolution sampling methods. Such studies also hold promise for identifying individual differences in affective dynamics and how these might relate to the onset and maintenance of mood and anxiety disorders.

To explore whether these affective processes are modulated by individual differences in social anxiety and depressive symptoms, we extended the winning model to include both main effects and interaction terms with each parameter for each symptom dimension. The main effects of both social anxiety and depression symptoms demonstrated considerable uncertainty, as evidenced by wide 95% highest density intervals— a pattern which is attributable to our specification of priors centered around zero. In contrast, the interaction effects yielded more robust findings. To the best of our knowledge, this is the first study to show that individuals with high trait social anxiety show stronger effects of experimentally-manipulated social feedback on their momentary mood and anxiety. Our findings align with those of an EMA study by Doorley et al. (2021) that showed that highly socially anxious individuals experienced bigger drops in their anxiety levels following a highly positive social event, compared to individuals with low social anxiety. These results contribute to the broader literature on the sensitivity to social threats and rewards in social anxiety. As previously mentioned, anxiety can be seen from an evolutionary perspective, as a mechanism for tracking potential threat and individuals high in social anxiety might be more sensitive to changes in social threat, as indicated by the higher sensitivity to social feedback. Consistent with this, a substantial body of research has demonstrated heightened sensitivity to social threat in socially anxious individuals (e.g. O'Connor et al. (2014); Cremers et al. (2015)). In contrast, findings on sensitivity to social rewards have been more mixed (Beltzer et al., 2023; Cremers et al., 2015; e.g. O'Connor et al., 2014). It is possible that social anxiety symptoms differentially influence momentary affect in response to social rewards and the capacity to learn from those rewards. That is, while socially anxious individuals may show a drop in their anxiety and an increase in their mood in response to positive social interactions, other processes, such as heightened self-focused attention or post-event processing, may inhibit their ability to encode or generalize these experiences over time (Brozovich & Heimberg, 2008; Clark & Wells, 1995). These findings also highlight a promising treatment opportunity: leveraging the therapeutic value of positive social surprises for the treatment of SAD. We provide experimental support for the idea that socially anxious individuals are affectively responsive to positive social feedback, indicating that interventions which amplify and consolidate these experiences, alongside

addressing maladaptive cognitive processes such as self-focused attention or post-event rumination, may enhance treatment outcomes.

We found that depressive symptoms were associated with heightened affective reactivity to negative PEs, specifically in mood. This finding diverges from that of Rutledge et al. (2017), that found no differences in neural or affective responses to RPEs between patients with Major Depressive Disorder (MDD) and controls. One key distinction is that our modelling approach allowed for asymmetric effects of positive and negative PEs, potentially uncovering individual differences that may have been obscured in their model which treat PEs as a single undifferentiated signal (Rutledge et al., 2014). Although, in an EMA study by Villano & Heller (2024), it was shown that depressive symptoms impaired emotional responses to positive, but not to negative PE. This discrepancy may reflect differences in the type of rewards used across studies; monetary rewards in Rutledge et al. (2017), exam grades in Villano & Heller (2024), and social feedback in ours. As such, heightened reactivity to negative PEs in our study may reflect a domain-specific sensitivity to social-evaluative threat rather than a general enhancement of PE-driven mood dynamics in depression. Furthermore, in contrast to existing literature emphasizing reduced reward sensitivity in depression, we found that depressive symptoms moderated the influence of social feedback on anxiety. These discrepancies, as well as the broader inconsistencies in the literature on reward learning in depression (Halahakoon et al., 2020; Kieslich et al., 2022), may be a manifestation of the substantial heterogeneity within depressive symptomatology (Fried, 2017). In other words, different symptoms may be associated with distinct abnormalities in how individuals learn from and emotionally respond to rewards and punishments, particularly in socially salient contexts. Future research could benefit from taking a symptom-specific approach, examining how features such as anhedonia, depressed mood, and self-criticism uniquely shape affective responses to different types of prediction errors.

Our study has several strengths. First, our modelling approach was grounded in preliminary analyses conducted on pilot data and then validated in a larger, preregistered sample, enhancing the robustness and reproducibility of our findings. Second, we sampled across three distinct recruitment strategies, including an online platform, schools, and the local community, capturing a diverse participant pool across a broad age range (14–45 years). This heterogeneity strengthens the generalizability of our results. Third, we simultaneously modelled momentary mood and

anxiety, allowing us to disentangle distinct computational mechanisms underlying different affective states. To our knowledge, this is the first study to extend computational models of mood to the social domain, and to explore how these dynamics relate to individual differences in social anxiety and depression. Finally, the experimental task was co-designed with a youth advisory group, ensuring clarity, engagement, relevance and developmental appropriateness for adolescent participants.

Nevertheless, several limitations warrant caution and highlight important directions for future research. Firstly, although the use of pre-determined social feedback allowed us to experimentally manipulate social prediction errors, this design may have limited the credibility of the social interactions, particularly since feedback was not contingent on participants' actual performance; however, the fact that participants' anxiety and mood were nevertheless influenced by the feedback suggests the manipulation retained psychological validity. Relatedly, while our model captured the effects of social prediction errors, the random trial structure prevented participants from forming stable expectations about any given partner and, in turn, limited our ability to model how they might have learned about or adapted to different social partners over time. Future research should explore how affective responses evolve during more naturalistic social interactions, ideally incorporating real-time belief updating and high-frequency affective sampling in daily life. Such work will be essential for testing the ecological validity of our findings and for identifying robust, computational markers of vulnerability to mood and anxiety disorders. Finally, while our sample spanned adolescence to adulthood, differences in age were confounded with recruitment method, limiting our ability to draw developmental inferences. Longitudinal and developmental studies will be essential to identify age effects on affective dynamics.

In sum, this study provides a computationally grounded account of how social experiences shape momentary affect, offering a framework that captures both shared and distinct mechanisms underlying mood and anxiety. Beyond contributing to theoretical models, these findings open new avenues for identifying early markers of risk and can inform more personalised approaches to treatment or early intervention.

Methods

Participants

We recruited 722 participants aged 18–25 through Prolific (www.prolific.com) across 21 pilot studies. Data from these pilots were used to refine the SPEAK task, inform the design of computational models based on preliminary mixed-effects model analyses, and conduct power calculations to estimate the required sample size.

We recruited 185 participants aged 14–45 years from three different sources: an online platform, UCL community, and secondary schools in the South East of England. Based on our preregistered power analyses using pilot data, which included parametric simulations for detecting the effect of social prediction errors on mood and anxiety and bootstrapping for Bayesian model comparison, both conducted to ensure at least 80% power, our final sample of 185 participants exceeds the required sample sizes ($n \approx 30$), indicating that the study is well powered. For participant demographic information see Table 1.

Online Participants. We recruited 106 online participants between the ages of 18-45 using Prolific (www.prolific.com). 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 on Prolific have been approved. We also required that participants had not taken part in any of our pilot studies. 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 flyers 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 SPEAK task. Participants were compensated with a £10 Love2shop or Amazon voucher for their time.

School Participants. We recruited 49 participants between the ages of 14-18 through schools (via physical flyers or via school newsletters). Schools were recruited either by direct contact or by advertising through the Anna Freud school network newsletter. Five participants were excluded due to poor engagement with the task, as evidenced by their video recordings. 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 SPEAK 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 as well as the donation of tablets or books.

Table 1: Summary demographics table by Recruitment Group

Recruitment Group	**N**	**Mean Age (± SD)**	**Female (Sex)**
Prolific Participants	98	26.7 (6.8)	61 (62.2%)
UCL Community Participants	43	21.7 (2.5)	40 (93.0%)
School Participants	44	16.4 (1.1)	27 (61.4%)

Questionnaires and the 'SPEAK' Task

The online questionnaires and the SPEAK task were created and hosted on Gorilla Experiment Builder (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-SR; 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 LSAS and CES-D and the low mood RCADS subscale.

After completing the mental health questionnaires and passing two attention checks, participants proceeded to the SPEAK 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 picture 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 harshness. 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. The picture stimuli were emotionally-neutral pictures depicting a scene (e.g. someone fixing a bike, a family cooking together etc.), chosen from free stock photo websites. After their description, 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 (PE^{N}; 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. PE^N feedback was set as either the exact histogram mean or one point above or below it. Note that these PE categories reflect our intended experimental manipulation, rather than participants' actual experienced PEs (see Computational Modelling section). The order of trials with each virtual player, the feedback received, the mood and anxiety rating prompts, and the picture stimuli were all randomised across participants and trials.

The SPEAK task was co-produced 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 credibility 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. A preview version of the questionnaires and task can be accessed here.

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 the SPEAK task and to isolate the unique contribution of each variable.

Calculation of social appraisal variables. Within this approach, the feedback or outcome amount, O_t, was defined as the percentage score received by the virtual player on trial t.

$$O_t = Score_t \qquad (3)$$

For all models except one, expectation, E_t, was defined as the participant-reported expectation on that trial. The only exception was model 2, in which E_t was defined as

$$E_t = \frac{1}{t-1}\sum_{i=1}^{t-1} O_i \qquad (4)$$

representing the average of all previous outcomes (O_i). When included in a time-decayed accumulator as shown in Equation 8 and Equation 9, this gives rise to a primacy bias; whereby earlier outcomes have a stronger influence on current mood or anxiety than more recent outcomes. This model was adapted from Keren et al. (2021) to fit the data from the SPEAK task.

Social PEs were defined as the difference between the actual outcome and expectation, E_t, on that trial.

$$PE_t = O_t - E_t \qquad (5)$$

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

$$\mathrm{PE}_{\mathrm{pos}}(t) = \begin{cases} \mathrm{PE}(t), & \text{if } \mathrm{PE}(t) > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

$$\mathrm{PE}_{\mathrm{neg}}(t) = \begin{cases} \mathrm{PE}(t), & \text{if } \mathrm{PE}(t) > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (7)$$

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 t for individual i was modelled as a normally distributed outcome:

$$\mathrm{Mood}_{it} \sim \mathcal{N}\left(\beta_{0i} + \beta_{1i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot \mathrm{Social\_Appraisal}_{ij}, \sigma_i\right) \qquad (8)$$

$$\mathrm{Anxiety}_{it} \sim \mathcal{N}\left(\beta_{0i} + \beta_{1i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot \mathrm{Social\_Appraisal}_{ij}, \sigma_i\right) \qquad (9)$$

Here, $Social\_Appraisal_{ij}$ denotes the value of the social appraisal on trial j. The parameter $\gamma_i \in (0,1)$ is an individual-specific forgetting factor that determines the influence of prior trials on self-reported affect. When, $\gamma = 0$ only the most recent trial contributes to affect, indicating no memory of past events. In contrast, $\gamma = 1$ 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:

$$\beta_{0i} = \mu_{\beta_0} + \epsilon_{\beta_{0i}} \qquad (10)$$

$$\beta_{1i} = \mu_{\beta_1} + \epsilon_{\beta_{1i}} \qquad (11)$$

$$\vdots$$

$$\beta_{ni} = \mu_{\beta_n} + \epsilon_{\beta_{ni}} \qquad (12)$$

$$\gamma_i = \Phi\left(\mu_\gamma + \epsilon_{\gamma_i}\right) \qquad (13)$$

$$\sigma_i = \exp\left(\mu_\sigma + \epsilon_{\sigma_i}\right) \qquad (14)$$

where $\epsilon_{\theta i} \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$ 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 (see Supplement).

Summary of Fitted Models. For our first aim, we fit and compared eight different computational models for mood and anxiety, respectively (see Table 2). 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 2: Summary table of models compared

Model	Expectation	Prediction Errors	Feedback
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 the 'SPEAK' 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 compute the Leave-One-Out Information Criterion (LOOIC) and the Watanabe-Akaike Information Criterion (WAIC), both of which estimate out-of-sample predictive fit by approximating leave-one-out cross-validation 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 (R^2) 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.

All outcome and predictor variables were z-scored prior to model fitting to place them on a common scale and facilitate interpretation of regression coefficients.

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, current trial social feedback, and a participant-specific moderator, Mi (e.g., LSAS-CA/LSAS-SR score for social anxiety, RCADS/CES-D score for depression):

$$\begin{aligned}
\text{Mood}_{it} &= \beta_{0i} + \beta_{1i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot \text{PE}_{\text{pos},j} + \beta_{2i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot \text{PE}_{\text{neg},j} + \beta_{3i} \cdot O_t + \beta_{4i} \cdot M_i \\
\text{Anxiety}_{it} &= \beta_{0i} + \beta_{1i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot \text{PE}_{\text{pos},j} + \beta_{2i} \cdot \sum_{j=1}^{t} \gamma_i^{t-j} \cdot \text{PE}_{\text{neg},j} + \beta_{3i} \cdot O_t + \beta_4 \cdot M_i \\
\beta_{0i} &= \mu_{\beta_0} + \beta_{0\times M} \cdot M_i + \epsilon_{\beta_{0i}} \\
\beta_{1i} &= \mu_{\beta_1} + \beta_{1\times M} \cdot M_i + \epsilon_{\beta_{1i}} \\
\beta_{2i} &= \mu_{\beta_2} + \beta_{2\times M} \cdot M_i + \epsilon_{\beta_{2i}} \\
\beta_{3i} &= \mu_{\beta_3} + \beta_{3\times M} \cdot M_i + \epsilon_{\beta_{3i}} \\
\gamma_i &= \Phi\left(\mu_\gamma + \beta_{\gamma\times M} \cdot M_i + \epsilon_{\gamma_i}\right) \\
\sigma_i &= \exp\left(\mu_\sigma + \beta_{\sigma\times M} \cdot M_i + \epsilon_{\sigma_i}\right)
\end{aligned} \tag{15}$$

where $\epsilon_{\theta i} \sim \mathcal{N}\left(0, \sigma_{\theta}^{2}\right)$ are individual-level deviations.

All predictors and outcomes were z-scored prior to model fitting. Because different measures were used for adults and adolescents, depression scores (CES-D in adults, RCADS in adolescents) and social anxiety scores (LSAS-SR in adults, LSAS-CA in adolescents) were each z-scored within age group to ensure comparability.

To compliment these analyses, we also computed the posterior distribution of Pearson correlations between each participant-level parameter estimate (e.g. social feedback, positive and negative PEs, forgetting factor) and participant's social anxiety or depression symptom scores. Correlations were computed for each posterior sample, providing a Bayesian estimate of the strength, 95% highest density interval (HDI), and probability of the association being in the observed direction for each parameter.

References

Beltzer, M. L., Daniel, K. E., Daros, A. R., & Teachman, B. A. (2023). Examining social reinforcement learning in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *80*, 101810. https://doi.org/10.1016/j.jbtep.2022.101810

Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2002). *Drug Use Disorders Identification Test*. https://doi.org/10.1037/t02890-000

Brozovich, F., & Heimberg, R. G. (2008). An analysis of post-event processing in social anxiety disorder. *Clinical Psychology Review*, *28*(6), 891–903. https://doi.org/10.1016/j.cpr.2008.01.002

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. In *Social phobia: Diagnosis, assessment, and treatment* (pp. 69–93). The Guilford Press.

Collishaw, S. (2015). Annual Research Review: Secular trends in child and adolescent mental health. *Journal of Child Psychology and Psychiatry*, *56*(3), 370–393. https://doi.org/https://doi.org/10.1111/jcpp.12372

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

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Cremers, H. R., Veer, I. M., Spinhoven, P., Rombouts, S. A. R. B., & Roelofs, K. (2015). Neural sensitivity to social reward and punishment anticipation in social anxiety disorder. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00439

Davidson, R. J. (1998). Affective Style and Affective Disorders: Perspectives from Affective Neuroscience. *Cognition and Emotion*, *12*(3), 307–330. https://doi.org/10.1080/026999398379628

Davidson, R. J. (2015). Comment: Affective Chronometry Has Come of Age. *Emotion Review*, *7*(4), 368–370. https://doi.org/10.1177/1754073915590844

Doorley, J. D., Goodman, F. R., Disabato, D. J., Kashdan, T. B., Weinstein, J. S., & Shackman, A. J. (2021). The momentary benefits of positive events for individuals with elevated social anxiety. *Emotion*, *21*(3), 595–606. https://doi.org/10.1037/emo0000725

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). Mood as Representation of Momentum. *Trends in Cognitive Sciences*, *20*(1), 15–24. https://doi.org/10.1016/j.tics.2015.07.010

Fan, R., Varol, O., Varamesh, A., Barron, A., Leemput, I. A. van de, Scheffer, M., & Bollen, J. (2019). The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nature Human Behaviour*, *3*(1), 92–100. https://doi.org/10.1038/s41562-018-0490-5

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

Forbes, L., & Bennett, D. (2023). The effect of reward prediction errors on subjective affect depends on outcome valence and decision context. *Emotion*, No Pagination Specified–No Pagination Specified. https://doi.org/10.1037/emo0001310

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

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Halahakoon, D. C., Kieslich, K., O'Driscoll, C., Nair, A., Lewis, G., & Roiser, J. P. (2020). Reward-processing behavior in depressed participants relative to healthy volunteers: A systematic review and meta-analysis. *JAMA Psychiatry*, *77*(12), 1286–1295. https://doi.org/10.1001/jamapsychiatry.2020.2139

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., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kieslich, K., Valton, V., & Roiser, J. P. (2022). *Pleasure, Reward Value, Prediction Error and Anhedonia* (D. A. Pizzagalli, Ed.; Vol. 58, pp. 281–304). Springer International Publishing. https://link.springer.com/10.1007/7854_2021_295

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

Leigh, E., & Clark, D. M. (2023). Internet-delivered therapist-assisted cognitive therapy for adolescent social anxiety disorder (OSCA): A randomised controlled trial addressing preliminary efficacy and mechanisms of action. *Journal of Child Psychology and Psychiatry*, *64*(1), 145–155. https://doi.org/10.1111/jcpp.13680

Marks, I. fM., & Nesse, R. M. (1994). Fear and fitness: An evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, *15*(5), 247–261. https://doi.org/10.1016/0162-3095(94)90002-7

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

O'Connor, E. J., Staiger, P. K., Kambouropoulos, N., & Smillie, L. D. (2014). Pathways to social anxiety: The role of reinforcement sensitivities and emotion regulation. *Psychiatry Research*, *220*(3), 915–920. https://doi.org/10.1016/j.psychres.2014.09.007

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., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J. Z., Ousdal, O. T., Prabhu, G., Dayan, P., Fonagy, P., & Dolan, R. J. (2017). Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression. *JAMA Psychiatry*, *74*(8), 790–797. https://doi.org/10.1001/jamapsychiatry.2017.1713

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

Shepperd, J. A., & Mcnulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science*, *13*(1), 85–88. https://doi.org/10.1111/1467-9280.00416

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

Villano, W. J., & Heller, A. S. (2024). Depression is associated with blunted affective responses to naturalistic reward prediction errors. *Psychological Medicine*, *54*(9), 1956–1964. https://doi.org/10.1017/S0033291724000047

Villano, W. J., Otto, A. R., Ezie, C. E. C., Gillis, R., & Heller, A. S. (2020). Temporal dynamics of real-world emotion are more strongly linked to prediction error than outcome. *Journal of Experimental Psychology: General*, *149*(9), 1755–1766. https://doi.org/10.1037/xge0000740

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*.

Acknowledgements

We gratefully acknowledge the Young Persons Advisory Group (YPAG) for their invaluable contribution to the co-production of the experimental task used in this study. Their insights and collaborative input were essential to the development of this research.

We thank Professor David M. Clark for his helpful comments on the manuscript.

This work was supported by the Wellcome Trust [226785/Z/22/Z].