**Methods**

**Participants**

Participants completed all measures during the tenth assessment wave of the Preschool Depression Study (PDS; see Supplemental Materials for full chronology of study). In this study, Participants were initially screened at preschool age for the study using the Preschool Feelings Checklist (PFC) (Luby et al., 2004) in order to recruit children with high and low symptoms of depression. Children were excluded if they presented with chronic illness, marked speech and/or language delays and/or neurologic or autism spectrum disorders. An additional 40 currently healthy children were added at school age. Of those, 118 participants completed the social feedback task. Of those, one participant was excluded due to technical errors during the social feedback task, and two participants were excluded due to intelligence quotients (IQ) below 70. Of those 115 remaining participants, 99 had completed and has useable data for the behavioral reward/loss task, 110 had completed a self-report measure of depression, and 115 had completed a self-report measure of social anxiety.

**Measures**

Depression severity

Depression severity was measured using the Child Depression Inventory–2 (CDI-2) (Kovacs, 1992) for participants less than 18 years-old and Beck Depression Inventory–II (BDI–II) (Dozois et al., 1998) for participants 18 years-old and older. The CDI-2 includes 28 items scored on a three point Likert scale. The BDI-II includes 21 items scored on a four point Likert scale. Both measures assess self-reported depression symptoms in the past 2 weeks and have excellent internal consistency (CDI-2: α=0.91; BDI-II: α=0.91) and test-retest reliability (CDI-2: *r*=0.89; BDI-II: *r*=0.93) (Beck et al., 1996; Dozois et al., 1998; Kovacs, 1992).

Social anxiety severity

Social anxiety severity was measured using the Social Interaction Anxiety Scale (SIAS-6) and Social Phobia Scale (SPS-6) (Peters et al., 2012)—a scale that combines the two scales and was developed as an abbreviated version of the full SIAS/SPS (Mattick & Clarke, 1998) using Item Response Theory modeling. This scale includes 6 items rated on a five point Likert scale from each of the SIAS-6 and SPS-6, comprising one total scale. The SIAS/SPS-6 assesses self-reported current symptoms of social anxiety/phobia (with no specific time frame given) and has excellent internal consistency (α > 0.90) and test-retest reliability (*r* > 0.91) (Mattick & Clarke, 1998; Osman et al., 1998).

Covariates

Covariates in multiple regression models included age at time of task, sex, race (Caucasian, African American, or Other), Hispanic ethnicity, and socioeconomic status (i.e., income-to-needs ratio).

**Procedure**

Social feedback ERP task

The Island Getaway task was used to measure brain responses to social feedback. For the task, participants are told they will be playing a game with 11 co-players in which they would be travelling in the Hawaiian Islands, and at each island, have to vote whether they want each co-player to continue on with them to the next island and then receive feedback on how co-players voted for them. Participants review information about each co-player (e.g., gender, location, personal preferences) and enter information that they are told is similarly reviewed by each co-player. Participants complete 6 rounds of “voting,” during which they vote whether to “keep” or “kick out” each co-player, and following each vote receive feedback as to whether that co-player voted to accept (“keep”) or reject (“kick out”) them (Figure 1). After each round, participants are told that one of the co-players had been sent out, and after completing the sixth round, participants are informed that they made it to the “Big Island.” Participants receive 51 trials of feedback: 25 acceptance and 25 rejection, and one randomly selected trial—a unique strength of this particular task (over tasks of only social rejection or exclusion) that yields reliable measurements of brain activity to social acceptance and rejection feedback. Co-players are randomly assigned a voting pattern for each participant such that 2 co-players reject the participant on most rounds, 2 co-players accept the participant on most rounds, and the remaining 7 co-players are equally likely to accept or reject the participant. Following the task, participants complete a brief post-task questionnaire assessing engagement in the task. This task has been validated in early adolescent, late adolescent, and young adult samples (Ethridge et al., 2017; Ethridge & Weinberg, 2018; Kujawa et al., 2017).

Behavioral reward bias and loss avoidance task

To measure reward bias and loss avoidance we used the Probabilistic Incentive Learning Task (PILT), specifically two modified versions (Heerey et al., 2008; Pizzagalli et al., 2005), here termed PILT-Positive and PILT-Negative, to assess reward bias and loss avoidance respectively. These tasks have been used in adult and pediatric samples (Heerey et al., 2008; Luking et al., 2016; Luking, Neiman, et al., 2015; Luking, Pagliaccio, et al., 2015; Pizzagalli et al., 2005, 2008). In each trial, participants perform a perceptual discrimination and indicate whether a long or short stimulus was briefly presented (Figure 1). For the PILT-P, a portion of correct responses receive gain feedback while, for the PILT-N, a portion of incorrect responses receive loss feedback. For both tasks, one of the two responses (i.e. short or long, termed the RICH response) is scheduled to receive three times the amount of feedback as the alternative (LEAN) response. This leads to preferentially selecting the RICH response in the PILT-P (reward bias) and avoiding the RICH response in the PILT-N (loss avoidance) (Luking, Neiman, et al., 2015; Luking, Pagliaccio, et al., 2015; Pizzagalli et al., 2005, 2008).

Participants complete 20 practice trials, followed by two blocks of ## trials each. Feedback was presented in a pseudorandom order, so that no more than three trials in a row could receive feedback. As in prior studies (Luking et al., 2016; Luking, Neiman, et al., 2015; Luking, Pagliaccio, et al., 2015; Pizzagalli et al., 2005), individual trials with reactions times (RT) faster than 150ms or slower than 2500ms were or great than or less than 3 standard deviations from the participant’s mean RT were excluded.

Response bias (log b) assesses behavioral responsiveness to feedback. More positive values on the PILT-P task indicate a greater propensity to select the RICH (i.e. rewarded) stimulus and more positive values on the PILT-N task indicate a greater propensity to select the LEAN (i.e. non-punished) stimulus.

EEG/ERP Acquisition and Processing

Continuous EEG was recorded using the BrainVision ActiChamp, 32 channel active channel amplifier system (BrainVision LLC, Morrisville, NC, USA). Electrodes were mounted in an elastic cap using a subset of the International 10/20 System sites (FP1, F3, F7, FC1, FC5, FT9, C3, T7, CP1, CP5, TP9, P3, P7, O1, Fz, Cz, Pz, Oz, FP2, F4, F8, FC2, FC6, FT10, C4, T8, CP2, CP6, P4, P8, TP10, O2), with a ground electrode located at FPz. The electrooculogram (EOG) generated from blinks and eye movements is recorded from five facial electrodes. The EEG is sampled at 500 Hz and all signals digitized on a computer.

All data was re-referenced to the average of Tp9 and Tp10 and band-pass filtered from 0.1 to 30 Hz. The EEG was corrected for EOG artifacts (Gratton et al., 1983) and physiological artifacts removed using an automatic procedure with a maximum allowed voltage step of 50 µV within a 400 ms interval length, maximum absolute difference between any two points of 175 µV, and a minimum allowed activity of 0.50 µV within a 100 ms interval length. The EEG is segmented into 1000 ms epochs, beginning 200 ms before and ending 800 ms after feedback onset. ERPs are averaged for acceptance and rejection feedback, and baseline corrected to activity 200 ms prior to feedback.

**Data Analysis**

PCA

Temporospatial principal component analysis (PCA) was conducted using the EP Toolkit (Dien, 2010b). A temporal PCA was conducted first, using a Promax rotation to rotate a simple structure in the temporal domain (Dien et al., 2007; Dien, 2010b, 2012) and including time points from each participant’s averaged data as variables and participants, recording sites, and trial types as observations. To identify factors accounting for substantial variance, a Scree plot (Cattell, 1966) was generated and a parallel test (Horn, 1965) conducted comparing a Scree of the dataset to that of a fully random dataset. We identified 15 temporal factor (TF) that accounted for a larger proportion of variance than the random dataset, and thus were retained. Next, a spatial PCA was conducted using Infomax rotation to rotate the spatial factors to independence (Dien, 2010a) and including all recordings sites as variables and participants, trial types, and temporal factor scores as observations. We identified 3 spatial factors (SF), resulting in 45 factor combinations accounting for a total 34.8% of the variance. Factor scores were converted to voltages and robust analysis of variance was conducted on factors that accounted for greater than 1% of variance to identify factors that meaningfully distinguish acceptance and rejection feedback, yielding 11 factors. Peak latency and electrode location were used to identify the RewP/FN component at temporal factor 2-spatial factor 1 (TF2/SF1). Data was imported from Brain Vision Analyzer to R (version 4.0.3; R Core Team, 2013) for all further computations.

The RewP was calculated as the mean activity 50ms before and 50ms after the peak latency (344ms, time window of 294-394ms)) at the peak electrode (FP2). In line with previous work and recommendations (Meyer et al., 2017), residual scores for the RewP response to acceptance accounting for the response to rejection were calculated, producing scores uncorrelated with the response to rejection. Likewise, residual scores for the FN response to rejection accounting for the response to acceptance were calculated, producing scores uncorrelated with the response to acceptance. This produces RewP and FN scores that are correlated but not inverses of one another and isolates mean amplitude in the ERP unique to acceptance or rejection. A non-difference score measure (e.g., mean amplitude to acceptance without accounting for activity to rejection) is confounded by activity throughout the brain unrelated to the acceptance stimulus.

Outliers detection and multiple imputation

Mahalanobis distance was calculated among all measures used in confirmatory analyses (i.e. reward positivity, feedback negativity, reward bias, loss avoidance, depression severity, social anxiety severity) and use to identify and exclude multivariate outliers with a threshold of *p*<0.001 (N=1). This resulted in a final sample of N=114.

Multiple imputation was conducted using the ‘MICE’ package (Buuren & Groothuis-Oudshoorn, 2011) in R, following a lack of significant evidence that the data were not missing completely at random (MCAR; ##stat##). Ten imputations were generated using predictive mean matching and the variables of interest (i.e. depression severity, social anxiety severity, reward positivity, feedback negativity, reward bias, and loss avoidance).

Multiple regression models

Following multiple imputation, measures were standardized in each of the 10 imputations. Multiple regression models were conducted and estimates were pooled across the imputations using the ‘miceadds’ package (Robitzsch & Grund, 2021).

Simultaneous multiple regression models were used to test all three primary hypotheses. That is, two models tested the first hypothesis. In one model, reward bias and loss avoidance were included as predictors, along with covariates, and the RewP as the outcome. The other model included the same predictors but included the FN as the outcome. These models allow for estimation of the amount of unique variance in the ERP component activity accounted for by reward bias and loss avoidance, controlling for covariates.

An identical structure was used to test the second and third hypotheses. That is, in models testing the second hypothesis, depression and social anxiety severity were included as predictors and the RewP as the outcome in the first model, and the FN as the outcome in the second model. In models testing the third hypothesis, depression and social anxiety severity were included as predictors and reward bias as the outcome in the first model, and loss avoidance as the outcome in the second model.

Structural equation model of hypotheses

A structural equation model (SEM) of all three hypotheses was modeled using the ‘lavaan’ package (CITE) to determine whether the hypotheses demonstrated an overall improvement in model fit over null model (i.e. one that assumed that all of the hypothesized relationships were null). By simultaneously testing the predicted associations between all variables, this model will account for within-subject shared variance across, for example, ERP components (i.e. RewP & FN), while testing its predicted associations with reward bias and loss avoidance and with the psychopathology measures. Model fit was assessed two ways. First, measures of absolute model fit—root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR)—were compared to established guidelines (RMSEA & SRMR < 0.08) to determine goodness of fit (68). Second, measures of relative model fit—comparative fit index (CFI), Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC)—were compared to a reduced model (with higher CFI and lower AIC and BIC indicating better model fit) (Hooper et al., 2008).

152 completed the behavioral reward task and 134 had useable data, 151 completed the behavioral loss task, 145 completed a self-report measure of depression, and 159 completed a self-report of social anxiety. Of those that completed each measure, 2 were excluded, 1 was excluded due to technical errors with the social feedback task.

T1, 3, 5, 6, 8, 10, 12, 14, 18, 20