Supporting Information

Rodman et al. 10.1073/pnas.1712398114

SI Methods

Age Analysis Approach. Analyses were performed by using a two-step process to query age effects and interrogate nonlinear patterns in the data. When computing linear or linear mixed-effects (LME) (for dependent variables with repeated measures) regression analyses, the *lm* and *lme* functions of the *nlme* package in R (1) were used to evaluate the statistical significance of standard age-related patterns of change (linear, quadratic, cubic). For models with repeated-measure dependent variables, subject was included as a random effect. Quadratic and cubic models contained lower-order age terms, and polynominal age terms were input as orthogonalized covariates of interest by using the *poly* function. As such, the linear model included a linear age predictor; the quadratic model included linear and quadratic age predictors; and the cubic model included linear, quadratic, and cubic age predictors.

When computing GAMs or GAMMs (for repeated-measures dependent variables) the *gam* and *gamm* functions of the *mgcv* package in R (2) were used to solve for patterns of age-dependent change using thin plate regression smoothing splines. This spline fitting technique does not require a priori assumptions (e.g., knots) and yields solutions that are penalized for complexity to avoid overfitting the data. All model comparisons were performed within-class of regression.

Participants. A total of 119 healthy individuals aged 10–23 were recruited from the local community. Participants' ethnic and racial diversity was representative of the local community (70.1% Caucasian, 13.2% African-American, 10.2% Asian, and 6.5% Hispanic). To ensure similarities in the demographic and socioeconomic characteristics across age, no more than 30% of adult participants were former or current students at Harvard University. All participants included in the final analyses reported fully believing the cover story. All online questionnaires were collected by using the secure online platform Qualtrics. Participants were recruited through online forums such as Craigslist, advertising in local newspapers, and flyers. Sample size was determined before data collection based upon sample sizes from a study using a similar task in a developmental population (3). A power analysis could not be conducted because the effect of this feedback-based task on changes in self-views (the key target dependent variable) has not been used in prior developmental work to our knowledge.

Exclusion criteria included history of a neurological disorder, current psychiatric disorder, and current use of psychotropic medication. A diagnostic clinical interview (4) was conducted during the study visit to confirm that participants did not meet criteria for current depression or anxiety disorder. Twelve participants were excluded from final analyses: five due their reported suspicion of the cover story; two due to insufficient variability in behavioral responses; two due to noncompliance resulting in incomplete data; two because experimenter-administered diagnostic interview indicated that they met criteria for a current neuropsychiatric disorder; and one for being under the influence of a mind-altering substance on the day of the study visit. The remaining 107 participants were included in key analyses.

For all dependent variables of interest, data were inspected for normality and outliers. Outliers were identified by using the Outlier Labeling Method (5) with a *g* multiplier of 2.2 (6) and were then Winsorized (7). The Winsor approach involves replacing outliers with the maximum (or minimum) value within the bounds of the outlier threshold. When completing the analysis examining the task-induced change in self-views, three additional participants were not included in this analysis because they did not complete the posttask self-esteem survey, resulting in 104 participants included in this analysis. Data from one participant aged 18.27 were identified as an outlier and Winsorized. When completing the analysis examining the task-induced change in likability ratings of peers, five additional participants were not included in this analysis because they did not complete the posttask ratings of peers, resulting in 102 participants included in this analysis. Data from one participant aged 11.21 were identified as an outlier and Winsorized.

Task Design. The cue phase (0–3,000 ms) presented a photograph of a peer, which remained visible for the duration of the trial (Fig. 1*B*, *Top*). Upon seeing the peer, participants predicted whether the peer had liked them with a button press of Yes or No. Following the participant's response, there was a brief delay (Fig. 1*B*, *Middle*; 1,000–4,000 ms) while the participant's prediction displayed on the left side of the screen, followed by delivery of feedback displayed on the right of the screen (Fig. 1*B*, *Bottom*; 2,000 ms), indicating whether the peer supposedly liked or disliked the participant (set at 50% acceptance, 50% rejection in pseudorandom order). Feedback was displayed in binary form as Yes or No, which participants were instructed mapped on to ratings in the top and bottom halves of the rating scale, respectively. If participants did not make a response in the allotted time during the cue phase, responses were coded as a "miss," and participants were not shown the peer feedback for that trial (1.7% of all trials).

Stimuli. We developed four sets of 160 face stimuli composed of headshot photographs of "peers" (ostensibly submitted by other participants) for four age groups (9–11, 12–14, 15–17, and 18+). Before conducting the study, we compiled a large database of candidate face stimuli depicting close-up headshots of individuals of various ages, sexes, and ethnicities from available stimulus sets (8–10) and from open-access image databases (e.g., Flickr, Pixabay, Wikimedia, Stockvault, and Freeimages).

These candidate images were then rated by a separate group of adult participants (n = 220) recruited through Amazon Mechanical Turk. These individuals rated images on: perceived age of the person (i.e., 9–11, 12–14, 15–17, or 18+), perceived ethnicity of the person (i.e., Caucasian, Asian or Pacific Islander, Hispanic, African American, Native American, or Other), perceived likability of the person [1 = (not at all) – 5 = (very much)], and whether the photograph looked like a realistic profile picture, the type of headshot peers may submit (Yes/No).

Based on these ratings, stimuli were sorted into four subsets of age-matched images to represent the other peers in the study. Each subset contained 160 images with a similar variety of ethnic backgrounds (no more than 60% Caucasian), equal portions of male and female images, images that the majority of raters considered to be realistic, and comparable ratings of likability. All faces were cropped, centered, resized, and presented against a black background.

The stimulus subsets were then validated to ensure comparability in likability using ratings from participants in the current study. Peer likability ratings were equal across age-specific stimulus subsets [$F(3, 106) = 0.704$, $P = 0.552$]. For the experimental task, face stimuli were pseudorandomized and counterbalanced so that each face delivered positive feedback and negative feedback at approximately equal rates across participants.

SI Results

Memory Control Analysis. LME regressions with memory accuracy as the dependent variable, subject as a random effect, feedback type (acceptance, rejection), age (linear, quadratic, cubic), and age interactions as predictors revealed that the null model without age predictors (AIC: −287.1) was superior to the cubic (AIC: −277.9), quadratic (AIC: −281.6), and linear (AIC: −285.4) models. This indicated that memory for social feedback did not vary as a function of age. To confirm the lack of age effects, we examined the significance level of age effects within the linear age model. There was no significant main effect of age (B = 0.262, $P = 0.132$) and no significant age by feedback type interaction (B = −0.231, $P = 0.345$; Fig. S5). Because there were no effects of age, more complex nonlinear models were not interrogated.

Age Group Analysis Based on Binned Data.

Age-binning procedure. We examined age-related differences across 2.5-y bins, an interval that allowed for equal distribution of participants across the entire sample: 10–12.5 y old (preadolescents; $n = 20$), 12.5–15 y old (early adolescents; $n = 23$), 15–17.5 y old (late adolescents; $n = 22$), 17.5–20 y old (emerging adults; $n = 19$), and 20–23 y old (young adults; $n = 23$).

Explicit: Prediction of peer feedback age group comparisons. A one-way ANOVA with proportion of predicted acceptance as a dependent variable and age group as a between-subjects factor revealed that predictions of being liked varied significantly with age [$F(4, 106) = 3.363$, $P = 0.013$; Fig. S1]. Post hoc between-group comparisons (corrected for multiple comparisons using Tukey's method) revealed that early adolescents (M = 47.17%, SE = 3.00%) predicted they would be liked significantly less often than emerging adults [M = 59.15%, SE = 2.91%; $t(40) = -2.833$, $P(\text{corr.}) = 0.031$] (where corr. is corrected) and young adults [M = 58.03%, SE = 2.81%; $t(44) = -2.642$, $P(\text{corr.}) = 0.045$]. All other age group comparisons were not significant [Ps(corr.) > 0.160].

Next, we examined the degree to which participants exhibited biased expectations of acceptance compared with the base rate of 50% (all participants received 50% acceptance and 50% rejection feedback across the task). Separate one-sample t tests of each age group revealed that this effect was carried by the adults [emerging adults: $t(18) = 3.146$, $P = 0.006$; young adults: $t(22) = 2.856$, $P = 0.009$], whereas the other age groups' predictions did not differ from 50% (all Ps > 0.356). In sum, while pre, early, and late-adolescents were objectively more accurate in their expectations of peer acceptance, they predicted being liked less frequently than emerging and young adults, who overestimated the extent to which they would be accepted by peers.

Implicit: Prediction response times age group comparisons. A one-way ANOVA with the prediction response time difference score (acceptance – rejection) as the dependent variable and age group as the between subjects factor revealed a significant effect of age [$F(4, 106) = 3.446$, $P = 0.011$; Fig. S2]. Post hoc between-group comparisons revealed that, compared with other age groups, early adolescents, the same group that anticipated being liked by peers the least often, were slowest to predict acceptance compared with rejection (M = 78.6 ms, SE = 28.5 ms). This pattern was significantly different from emerging adults, who showed an opposite pattern of speeding to predict acceptance [M = −58.2 ms, SE = 31.5 ms; $t(40) = 3.221$, $P(\text{corr.}) = 0.008$]. No other group comparisons reached significance [all Ps(corr.) > 0.103].

Results of one-sample t tests within age group (against the null hypothesis of zero response time bias) indicated that early adolescents were significantly slower when predicting acceptance compared with rejection [$t(23) = 2.759$, $P = 0.011$]. By contrast, emerging adults were somewhat slower when predicting rejection relative to acceptance [$t(18) = -1.846$, marginal $P = 0.081$]. All other participant groups did not exhibit differences in response time for acceptance compared with rejection (all Ps > 0.218). In sum, early adolescents exhibited an internal heuristic more consistent with expecting rejection from others, whereas emerging adults demonstrated an internal heuristic more consistent with expecting others to accept them.

Changes in views of self age group comparisons. A one-way ANOVA with the self-esteem percent change score as the dependent variable and age group as the between-subjects factor indicated that there were significant age differences in task-induced changes in self-views [$F(4, 103) = 3.382$, $P = 0.012$; Fig. S3]. Post hoc between-group comparisons revealed that early adolescents experienced a unique drop in self-views (M = −5.08%, SE = 2.48%) that differed significantly from the rise experienced by late adolescents [M = 3.54%, SE = 1.95%; $t(41) = -2.747$, $P(\text{corr.}) = 0.042$], emerging adults [M = 4.60%, SE = 2.83%; $t(37) = -2.303$, $P(\text{corr.}) = 0.022$], and young adults [M = 3.86%, SE = 2.03%; $t(42) = -2.808$, $P(\text{corr.}) = 0.029$]. The rise in self-views experienced by all other participant groups was equivalent [all Ps(corr.) > 0.416].

Next, a set of one-sample t tests evaluated the magnitude of these shifts in self-views relative to a null hypothesis of zero change. Within each age group, findings revealed that the drop in self-views experienced by early adolescents was marginally significant [$t(20) = -2.049$, $P = 0.054$], as was the rise in self-views observed in late adolescents and young adults [$t(21) = 1.816$, marginal $P = 0.084$; $t(22) = 1.899$, marginal $P = 0.071$, respectively]. There was not a significant task-induced change in self-views exhibited by any other age groups (all Ps > 0.121). Together, these findings suggest that the experience of mixed social evaluative feedback differentially impacted self-views across age, wherein self-enhancement emerged late during adolescence.

Changes in likability ratings of peers age group comparisons. Key analyses tested whether the tendency to increase liking after acceptance and decrease liking after rejection varied with age. A repeated-measures ANOVA testing for age effects on change in likability ratings following acceptance and rejection revealed a significant interaction [$F(4, 98) = 2.670$, $P = 0.037$; Fig. S4]. Pairwise comparisons within each age group revealed that emerging adults ($M_{\Delta LIKE} = -0.69$, $SE_{\Delta LIKE} = 1.97$; $M_{\Delta DISLIKE} = -4.47$, $SE_{\Delta DISLIKE} = 2.08$) and young adults ($M_{\Delta LIKE} = 3.10$, $SE_{\Delta LIKE} = 1.57$; $M_{\Delta DISLIKE} = -4.36$, $SE_{\Delta DISLIKE} = 1.96$) adjusted likability ratings of peers upward or downward in accordance with whether they were liked or disliked [$t(17) = 1.935$, marginal $P = 0.070$; $t(22) = 6.566$, $P < 0.001$, respectively]. All other groups did not alter likability ratings as a function of whether they were liked or disliked (all Ps > 0.094).

One-sample t tests were computed within each age group for acceptance and rejection feedback, separately, to compare the magnitude of these shifts in likability ratings relative to a null hypothesis of zero change. Findings revealed that emerging and young adults showed a significant decrease in likability rating of peers who rejected them [$t(17) = -2.151$, $P = 0.046$; $t(22) = -2.220$, $P = 0.037$, respectively], whereas only young adults showed an increase in likability ratings of peers that accepted them [$t(22) = 1.970$, marginal $P = 0.062$]. All other groups did not demonstrate significant changes in ratings of the peers (all Ps > 0.127). Together, findings showed that being accepted or rejected by peers impacted how emerging and young adults viewed their peers, while other participant groups retained consistent evaluations of peers following social feedback.

Memory control analysis age group comparisons. A repeated-measures ANOVA with the memory accuracy as the outcome variable, feedback type (acceptance, rejection) as a within-subjects factor, and age group as the between-subjects factor revealed that the interaction between feedback type and age group was not significant [$F(4, 102) = 0.515$, $P = 0.725$; Fig. S5]. The main effect of age was also not significant [$F(4, 102) = 0.360$, $P = 0.836$].

1. Pinheiro J, Bates D, DebRoy S, Sarker D (2014) nlme: Linear and Nonlinear Mixed Effects Models. Available at https://cran.r-project.org/web/packages/nlme/index.html. Accessed October 9, 2017.
2. Wood SN (2017) *Generalized Additive Models: An Introduction with R* (CRC, Boca Raton, FL), 2nd Ed.
3. Gunther Moor B, van Leijenhorst L, Rombouts SARB, Crone EA, Van der Molen MW (2010) Do you like me? Neural correlates of social evaluation and developmental trajectories. *Soc Neurosci* 5:461–482.
4. Sheehan DV, et al. (1998) MINI-Mini International neuropsychiatric interview–English version 5.0.0-DSM-IV. *J Clin Psychiatry* 59:34–57.
5. Hoaglin DC, Iglewicz B, Tukey JW (1986) Performance of some resistant rules for outlier labeling. *J Am Stat Assoc* 81:991–999.
6. Hoaglin DC, Iglewicz B (1987) Fine-tuning some resistant rules for outlier labeling. *J Am Stat Assoc* 82:1147–1149.
7. Dixon WJ (1960) Simplified estimation from censored normal samples. *Ann Math Stat* 31:385–391.
8. Egger HL, et al. (2011) The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): A new set of children's facial emotion stimuli. *Int J Methods Psychiatr Res* 20: 145–156.
9. Jones RM, et al. (2011) Behavioral and neural properties of social reinforcement learning. *J Neurosci* 31:13039–13045.
10. Tottenham N, et al. (2009) The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Res* 168:242–249.

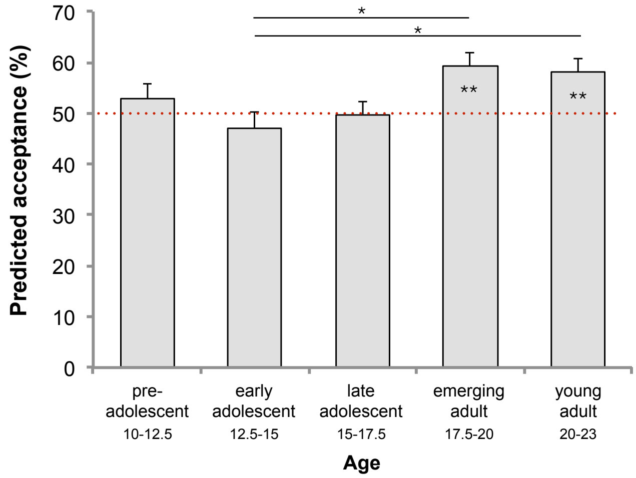

Fig. S1. Relative to adolescents, young adults overestimated how much they would be liked by peers. Graph shows average percentage of acceptance predictions for each age group. Error bars indicate SEMs. Red dotted line denotes actual rate of acceptance (50%). $^*P < 0.05$; $^{**}P < 0.01$.

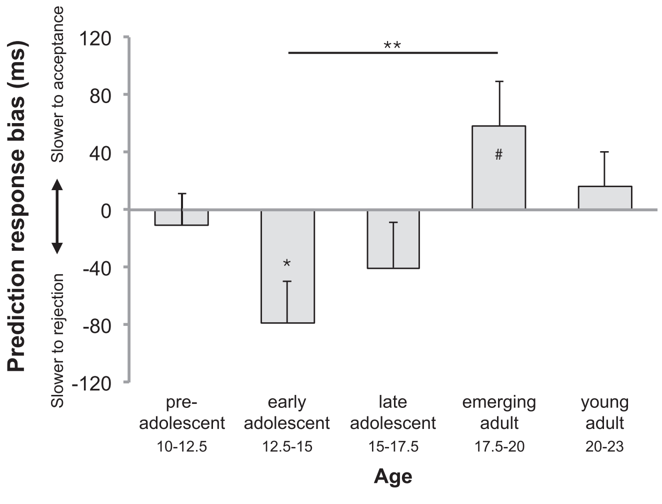

Fig. S2. Response time bias for predicting rejection vs. acceptance differs with age. Graph shows difference score of prediction response time during acceptance vs. rejection for each age group. Positive values indicate slowing when predicting rejection, whereas negative values indicate slowing when predicting acceptance. Error bars indicate SEM. $^\#P < 0.085$; $^*P < 0.05$; $^{**}P < 0.01$.

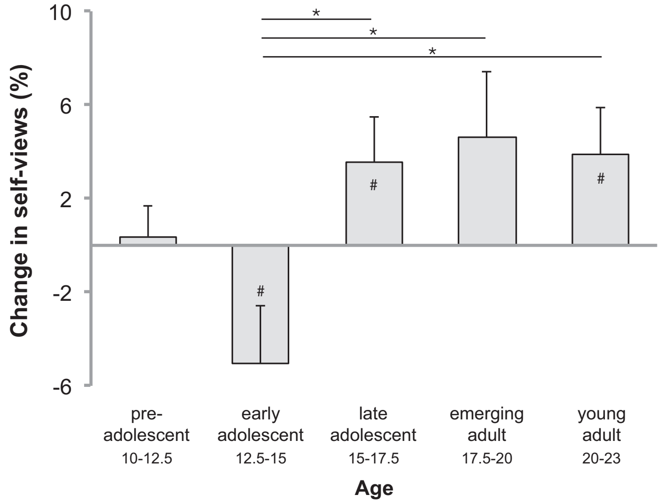

Fig. S3. Task-induced changes in self-views differ across development. Graph shows the percent change in self-esteem score pretask vs. posttask. Positive values indicate a boost in self-views, whereas negative values indicate a reduction in self-views. Error bars indicate SEM. #$P < 0.085$; *$P < 0.05$.

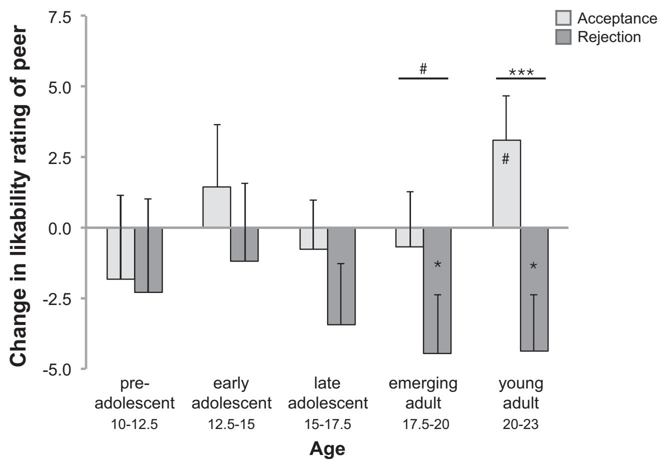

Fig. S4. Feedback-specific changes in likability ratings of peers present in emerging and young adults only. Graph shows the differences score in peer likability rating pretask vs. posttask as a function of feedback type. Light gray bars indicate change in likability rating following acceptance, and dark gray bars indicate change in likability rating following rejection. Positive values indicate an upgrade in peer rating, whereas negative values indicate a downgrade in peer rating. Error bars indicate SEM. #$P < 0.085$; *$P < 0.05$; ***$P < 0.001$.

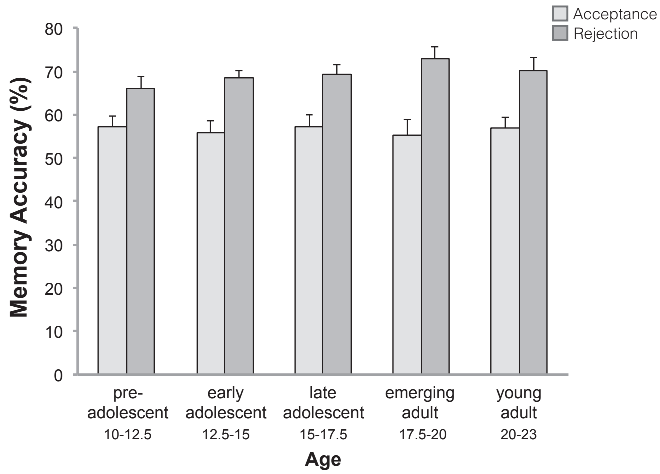

Fig. S5. Memory retention control analysis. Memory accuracy for feedback is equivalent across age. Light gray bars indicate accuracy for acceptance and dark gray bars indicate accuracy for rejection. Error bars indicate SEM.

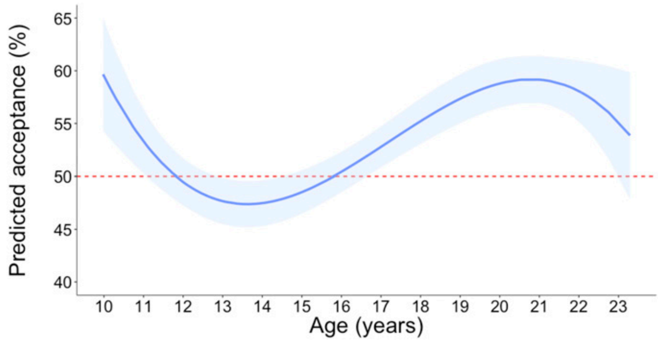

Fig. S6. Relative to adolescents, young adults overestimated how much they would be liked by peers. Graph shows a blue fit line of predicted values of acceptance expectancy (percent of trials predicted acceptance) based on the traditional cubic regression model. Blue shading indicates SEM. Red dotted line denotes actual rate of acceptance (50%).

Table S1. Self-views are comparable to normed levels for each age group

Age	*n*	$Mean_{SAMPLE}$	SE_{SAMPLE}	$Mean_{NORM}$
10–14	42	3.07	0.07	3.09
15–17	25	3.20	0.12	2.98
18–23	39	3.13	0.11	3.19

Age groups binned by version of SPP administered. Within each age group, participants scored a score roughly consistent with normed means (one-sample *t* test against normed mean for each scale: all *P*s > 0.078).