Title

Altered Reinforcement Learning from Reward and Punishment in Anorexia Nervosa: Evidence from Computational Modeling

Journal

Journal of the International Neuropsychological Society, 28(10)

ISSN

1355-6177

Authors

Wierenga, Christina E
Reilly, Erin
Bischoff-Grethe, Amanda
et al.

Publication Date

2022-11-01

DOI

10.1017/s1355617721001326

Peer reviewed

HHS Public Access

Author manuscript

J Int Neuropsychol Soc. Author manuscript; available in PMC 2022 November 01.

Published in final edited form as:

J Int Neuropsychol Soc. 2022 November ; 28(10): 1003–1015. doi:10.1017/S1355617721001326.

Altered Reinforcement Learning from Reward and Punishment in Anorexia Nervosa: Evidence from Computational Modeling

Christina E. Wierenga[1,*], **Erin Reilly**[2], **Amanda Bischoff-Grethe**[1], **Walter H. Kaye**[1], **Gregory G. Brown**[1]

[1]University of California, San Diego, CA, USA

[2]Hofstra University, Hempstead, NY, USA

Abstract

Objectives: Anorexia nervosa (AN) is associated with altered sensitivity to reward and punishment. Few studies have investigated whether this results in aberrant learning. The ability to learn from rewarding and aversive experiences is essential for flexibly adapting to changing environments, yet individuals with AN tend to demonstrate cognitive inflexibility, difficulty set-shifting and altered decision-making. Deficient reinforcement learning may contribute to repeated engagement in maladaptive behavior.

Methods: This study investigated learning in AN using a probabilistic associative learning task that separated learning of stimuli via reward from learning via punishment. Forty-two individuals with Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 restricting-type AN were compared to 38 healthy controls (HCs). We applied computational models of reinforcement learning to assess group differences in learning, thought to be driven by violations in expectations, or prediction errors (PEs). Linear regression analyses examined whether learning parameters predicted BMI at discharge.

Results: AN had lower learning rates than HC following both positive and negative PE ($p < .02$), and were less likely to exploit what they had learned. Negative PE on punishment trials predicted lower discharge BMI ($p < .001$), suggesting individuals with more negative expectancies about avoiding punishment had the poorest outcome.

Conclusions: This is the first study to show lower rates of learning in AN following both positive and negative outcomes, with worse punishment learning predicting less weight gain. An inability to modify expectations about avoiding punishment might explain persistence of

[*]Correspondence and reprint requests to: Christina E. Wierenga, Ph.D., Professor of Psychiatry, UCSD Eating Disorder Research and Treatment Program UCSD Department of Psychiatry, University of California, Chancellor Park, 4510 Executive Dr., Suite 315, San Diego, CA, 92121, USA. cwierenga@ucsd.edu.

SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit https://doi.org/10.1017/S1355617721001326

CONFLICT OF INTEREST

None of the authors have conflicts of interest to disclose.

ETHICAL STANDARDS

The study was approved by the Institutional Review Board of the University of California, San Diego, research was completed in accordance with the Helsinki Declaration, and all participants gave written informed consent and received a stipend.

restricted eating despite negative consequences, and suggests that treatments that modify negative expectancy might be effective in reducing food avoidance in AN.

Keywords

Eating disorders; prediction error; operant learning; decision-making; cognition; probabilistic associative learning

INTRODUCTION

Anorexia nervosa (AN) is a serious eating disorder characterized by severe food avoidance and weight loss, an intense fear of gaining weight, and a distorted experience of one's body (American Psychiatric Association, 2000). It is well known that individuals with AN tend to be cognitively inflexible and have impaired set-shifting, which may contribute to the high rates of chronicity and death (Papadopoulos, Ekbom, Brandt, & Ekselius, 2009; Roberts, Tchanturia, Stahl, Southgate, & Treasure, 2007; Roberts, Tchanturia, & Treasure, 2010; Tchanturia et al., 2012; Wu et al., 2014). Persistent dietary restriction despite negative consequences and evidence of altered reward and punishment sensitivity in AN (Bischoff-Grethe et al., 2013; Glashouwer, Bloot, Veensra, Franken, & de Jong, 2014; Harrison, O'Brien, Lopez, & Treasure, 2010; Harrison, Treasure, & Smillie, 2011; Jappe et al., 2011; Matton, Goossens, Braet, & Vervaet, 2013) raise the question of whether impaired learning from reward and loss might contribute to repeated engagement in maladaptive behavior and illness maintenance.

Dysfunction of reward processing in AN is well documented, with reduced subjective reward sensitivity and decreased limbic-striatal neural response to rewarding stimuli such as food or money (Brooks, Rask-Andersen, Benedict, & Schioth, 2012; Fladung, Schulze, Scholl, Bauer, & Gron, 2013; Jappe et al., 2011; Keating, Tilbrook, Rossell, Enticott, & Fitzgerald, 2012; O'Hara, Schmidt, & Campbell, 2015; Oberndorfer et al., 2013; Wierenga et al., 2014; Wu et al., 2016). Emerging evidence suggests processing of aversive stimuli may also be disrupted in AN; individuals with AN demonstrate elevated harm avoidance, intolerance of uncertainty, anxiety, and oversensitivity to punishment (Glashouwer et al., 2014; Harrison et al., 2010; Harrison et al., 2011; Jappe et al., 2011; Matton et al., 2013), which may contribute to an altered response to negative feedback or a bias to avoid outcomes perceived as aversive (Kaye et al., 2015). Neuroimaging studies support a neural dysfunction to loss, with an exaggerated (Bischoff-Grethe et al., 2013) or undifferentiated (Wagner et al., 2007) striatal response to monetary losses compared to wins and decreased response to aversive taste (Monteleone et al., 2017). However, much of the existing work in AN has focused on responsivity to reward and punishment, with less attention to learning from both reward and punishment (Bernardoni et al., 2018; Foerde & Steinglass, 2017).

The core idea of reinforcement learning is that the rate of learning is driven by violations of expectations, or prediction errors (PEs), which are operationalized as the received outcome minus the expected outcome, and are markers of dopamine activity (Pearce & Hall, 1980; Rescorla and Wagner 1972; Sutton & Barto, 2018). Learning from experience occurs through updating expectations about the outcome in proportion to PE, so that the

expected outcome converges to the actual outcome. The majority of studies of learning in AN have focused on passive Pavlovian conditioning (Schaefer & Steinglass, 2021), with evidence of elevated reward PE signals in the ventral striatum and orbitofrontal cortex in ill and remitted AN (GK Frank, Collier, Shott, & O'Reilly, 2016; GK Frank et al., 2012). However, Pavlovian tasks have demonstrated poor behavioral profiles (National Institute of Mental Health, 2016). Given the importance of choice behavior and decision-making in AN, instrumental response-outcome learning may be more relevant to psychopathology. Limited behavioral data (i.e., Acquired Equivalence Task) suggest reduced reward reinforcement learning in AN (Foerde & Steinglass, 2017; Shott et al., 2012).

To probe the influence of rewarding and punishing outcomes on instrumental reinforcement learning, we employed a well-studied two-choice feedback-based probabilistic associative learning task (PALT) that relies on the contingency between a participant's response and outcome (i.e., whether or not they won or lost points) to facilitate learning (i.e., to select the optimal reward-based stimuli and avoid the nonoptimal punishment-based stimuli) (Bodi et al., 2009; Herzallah et al., 2017; Herzallah et al., 2013; Mattfeld, Gluck, & Stark, 2011; Myers et al., 2013). The PALT is sensitive to dopaminergic medication effects on reward and punishment processing in Parkinson's disease (Bodi et al., 2009), has been applied to several psychiatric disorders (i.e., substance use, post-traumatic stress, depression (Beylergil et al., 2017; Herzallah et al., 2017; Myers et al., 2013), and corresponds to functional specialization within the striatum for reward and punishment PE estimates (Mattfeld et al., 2011). Moreover, research over the past two decades has shown that the direction and magnitude of PE may be a marker of altered dopaminergic activity in AN (Glimcher, 2011; Schultz, Dayan, & Montague, 1997; Schultz, 2016; Steinberg et al., 2013).

Given the link between PE and reinforcement learning, it is tempting to infer group or individual differences in PE from observable reinforcement learning scores. Such an inference would be valid only if the observed scores were unidimensional and reflected PE-based learning. However, if PALT performance involved multiple processes, group or individual differences in the observed scores would be challenging to interpret because the differences might be due to any of the several processes that underlie the task (Sojitra, Lerner, Petok, & Gluck, 2018; Strauss & Smith, 2009). Before comparing AN and healthy control (HC) participants, we investigated the multidimensionality of data derived from the PALT by comparing the fits of three computational reinforcement learning models.

All of these models assumed that when a stimulus is presented, participants choose between two alternatives based on unobserved choice values that reflect the participant's expectancy of obtaining a favorable outcome (See Supplement). Once a choice is made, the expectancy value associated with the choice made is updated based on the PE and PE learning rates, represented by the parameter η (Figure 1). In expectancy value-based learning models of this type, the difference between the expectancy values associated with the two-choice alternatives is multiplied by a logistic regression weight, represented by the parameter β, to turn the value difference into a probability of choosing a particular alternative (Gershman, 2016); Supplement – Equation 1; Figure 1). Although the logistic regression weight has been called inverse temperature in some applications (Daw, 2011), it has been described as an explore-exploit parameter in the psychology literature and reflects how decisively

participants make choices based on small differences in the expectancy values (Gershman, 2016; Moustafa, Gluck, Herzallah, & Myers, 2015).

As shown by Shultz (Schultz, 2016), positive and negative PEs differentially effect dopaminergic activity. Because differential levels of dopaminergic activity influence amount of PE learning (Steinberg et al., 2013), positive and negative PE might be associated with different PE learning rates. All models discussed in this paper assume that separate learning parameters differentially update expectancy values depending on the positive or negative valence of the PE (Gershman, 2016). In particular, the No Bias model is composed of the explore-exploit parameter, β, and two learning rate parameters, one to update expectancy values when PE is positive, η_p, the other when it is negative, η_n.

The No Bias model assumes that the first choice made to a stimulus is unbiased. However, global choice biases, the tendency to choose one alternative over another regardless of previous outcomes, and choice inertia bias, the tendency to repeat choices, are commonly reported in the choice literature (Fritsche, Mostert, & de Lange, 2017; Garcia-Perez & Alcala-Quintana, 2013; Gold & Ding, 2013; Linares, Aguilar-Lleyda, & Lopez-Moliner, 2019; Morgan, Dillenburger, Raphael, & Solomon, 2012). It is during experimental conditions leading to uncertainty that choice biases are most likely to be observed (Morgan et al., 2012; Urai, Braun, & Donner, 2017). When a stimulus is first presented on the PALT, participants are doubly uncertain, neither knowing whether the trial is a reward or punishment trial nor knowing which category to choose. Given this uncertainty, initial choice biases might be due to a global choice bias or to a choice history bias – the latter occurring on the initial presentation of subsequent stimuli after the first PALT stimulus is presented. If choice biases occur on the PALT, they would be unobserved processes that would obscure the use of observed scores as markers of PE learning. In the First Choice Bias model, we modeled the impact of choice biases on the expectancy value of a choice when a stimulus is first presented, which is when uncertainty is likely maximal. This model included a separately estimated bias parameter, bias(s$_j$), for each of the four stimuli, s$_j$, presented on a trial set in addition to the explore-exploit parameter, β, and the two learning rate parameters, η_p and η_n. The First Choice Bias (Singlet) model constrained estimates of the four bias parameters to be equal to a single estimated value.

Considering the importance of biases in accounting for choice performance, we predicted that the First Choice Bias model would provide a better fit to the data than would the Base model. Once the best fitting model was chosen, we tested the hypothesis that individuals with AN would demonstrate deficient reinforcement learning as evidenced by worse optimal response accuracy on reward and punishment trials and/or poorer learning rates, $\eta_{p/n}$, associated with positive and negative PEs compared to HCs. Moreover, within AN, differences between accuracy on reward and punishment trials or positive and negative PEs would indicate differential sensitivity to learning from rewarding or disappointing outcomes. Exploratory analyses examined associations between learning rates, size of PEs and AN symptom severity and clinical outcome.

METHOD

Participants

Forty-two individuals meeting criteria for DSM-5 restricting-type AN (4 also endorsed purging; mean age = 22.8, range = 16–60) were compared to 38 HC volunteers (mean age = 21.6, range =15–32; Table 1). Individuals with AN were recruited from the University of California, San Diego Eating Disorders Treatment and Research outpatient Partial Hospitalization Program (PHP). The PHP uses a blend of family-based treatment and dialectical behavior therapy adapted for intensive treatment settings. Patients received treatment 6 to 10 h/day, 6 days/week, including individual, family, group, and multi-family therapy, nutritional counseling, psychiatric care, and medical monitoring (Brown et al., 2018; Reilly et al., 2020). AN diagnosis was determined by semi-structured interview performed by program psychiatrists at treatment admission according to 2010 draft criteria for the DSM-5 (Hebebrand & Bulik, 2011) and included atypical and partially remitted AN (BMI range: 14.5–23.8 kg/m^2). HCs were recruited from the San Diego community and did not have any eating disorder symptomatology or Axis I psychiatric disorder based on a modified version of the Structured Clinical Interview for *DSM-IV-TR* Module H (First, Spitzer, Gibbon, & Williams, 2002) and the Mini International Neuropsychiatric Interview (Sheehan et al., 1998). See Supplement for additional exclusion criteria.

Procedure

AN participants completed the PALT on average 19.8 days (SD = 19.9) after treatment admission. Weight and height, measured via digital scale and stadiometer, were obtained at admission, within two days of PALT completion, and at discharge for AN, and during the task visit for HC. Self-report questionnaires to assess anxiety, depression and temperament traits common in AN (e.g., reward/punishment sensitivity, inhibition, harm avoidance) that might relate to learning behavior (Table 1) were completed within 16.1 days (SD = 18.9) of the PALT in AN (Harrison, Treasure, & Smillie, 2011; Jappe et al., 2011; Wagner et al., 2006). The study was approved by the Institutional Review Board of the University of California, San Diego, research was completed in accordance with the Helsinki Declaration, and all participants gave written informed consent and received a stipend.

Probabilistic Associative Learning Task

The PALT (Figure 2) involves receiving 25 points when choosing the optimal response on reward trials, but losing 25 points when choosing the nonoptimal response on punishment trials (Bodi et al., 2009; Mattfeld et al., 2011; Myers et al., 2013). On each trial, participants saw one of four stimulus images and were prompted to decide whether it was associated with one of two categories “A” or “B”, corresponding to different response keys. Two images were randomly assigned to be “reward” stimuli in that selection of the optimal category typically produced feedback and a gain of 25 points, whereas selection of the nonoptimal category typically produced no gain of points. The remaining two images were “punishment” stimuli in that selection of the nonoptimal category typically produced feedback and a loss of 25 points, whereas selection of the optimal category typically produced no loss of points. Reward-learning trials and punishment learning trials were intermixed within the task with a favorable outcome associated with a gain on reward

trials and the avoidance of loss on punishment trials. Unfavorable outcomes led to no change in points on reward trials and a loss of 25 points on punishment trials. The participant's cumulative point tally was shown at the bottom of the screen on each trial and was initialized to 500 points at the start of the experiment. As done in prior studies (Bodi et al., 2009; Mattfeld et al., 2011), two task sets were administered, each with a different set of pictures to increase the number of trials during which participants were actively learning new associations. The order of stimulus sets was counterbalanced across participants. Each set contained 160 trials, divided into four 40-trial blocks. Within each block, each stimulus appeared 10 times; 8 times the optimal response was associated with the more favorable outcome, whereas two times the nonoptimal response was associated with the more favorable outcome. For each participant, trial order was randomized within a block. Trials lasted until the participant responded and were separated by a 2s interval, during which time the screen was blank. On each trial, the computer recorded whether the participant made the optimal response, regardless of the actual outcome on that trial. The task took about 30 min to complete. The experiment was administered on a MacBook Pro, programmed in MatLab version R2016B.

Computational Reinforcement Learning Models

Like Confirmatory Factor Analysis, computational models of cognitive processes embody assumptions about a model's architecture and parameters that determine how observed data are related to latent processes. Whereas the assumptions fix the architecture of a model, varying the model's parameters can fine-tune the model's functioning (Farrell & Lewandowsky, 2018). Parameters estimated for each of the three models are listed in Table 2 and discussed in more detail in the caption of Figure 1 and in Supplemental Materials. To operationalize PE size, outcome was coded 1 for gains on reward trials, −1 for loss on punishment trials, and 0 for no change in points. Successful learning drives the expectancy values toward gains, coded 1, on reward trials and toward avoidance of loss, coded 0, on punishment trials. The No Bias model allowed positive and negative PE learning rate parameters, η_p and η_n, and the explore-exploit parameter, β, to vary and set initial expectancy values to zero. The First Choice Bias model (Figure 1) allowed β, η_p and η_n to vary, but also included four parameters that determined the initial expectancy values of choices made to each of the four stimuli in order to account for choice biases. Given how expectancy values are updated, the impact of these biases propagates to subsequent trials. The First Choice Bias (Singlet) model set the four bias parameters to the same estimated value. The full First Choice Bias model was selected as the best fitting model as assessed by deviance information criterion weights (see Supplement).

Parameter estimation—We used the R routine rjags to generate Bayesian estimates of model parameters based on fits to trial by trial optimal response data for each stimulus (Plummer, 2017). See Supplement for details and model sensitivity analysis. The predicted block means for reward and punishment trials based on parameter estimates for the best fitting model are presented in Figure 3.

Statistical Analysis

Behavioral performance—Choice behavior was analyzed using a repeated measures analysis of variance (rmANOVA) on optimal response accuracy with Group as a between subjects effect and Block and Set as within subject effects, separately for reward trials and punishment trials.

Model-generated parameters—Analyses were performed separately for reward and punishment trials. To compare groups on learning rate parameters, we performed a rmANOVA with Group as a between effect and Set and PE learning rates (η_p, η_n) as within effects. We also performed a Group x Set rmANOVA to investigate group differences in the β parameter. To investigate the bias parameters, we averaged the two bias values for reward stimuli and the two bias values for punishment stimuli, then performed a rmANOVA involving Group × Set. To more completely examine group differences in level of learning from a PE perspective, we averaged the size of PEs over trials separating values by PE type (positive or negative) within reward and punishment trials for each set (e.g., mean negative PE for punishment trials on set 1) and submitted these means to Group × Set × PE type rmANOVAs.

Exploratory clinical associations—To examine whether standard clinical assessments are associated with learning in AN, Pearson correlational analyses examined relationships between 14 reinforcement learning model values (for each set: η_p, η_n, positive and negative PEs for each trial type, and β) and 9 AN clinical measures (age, admission BMI, EDE-Q Global score, TCI Harm Avoidance, TCI Novelty Seeking, BIS/BAS, SPSRQ, STAI, BDI) at time of study. To examine associations with treatment outcome, reinforcement learning model values were explored as predictors of BMI at discharge using hierarchical linear regression analyses, controlling for BMI at treatment admission, length of treatment, and medication status. The hierarchical linear regression analysis was repeated using each self-reported clinical measure as a predictor. Bonferroni correction for multiple comparisons was used to determine a family-wise p-value for the 14 learning model values (.004) and the 9 clinical measures (.006) assuming $p = .05$ for each test.

Sensitivity analyses—To examine the potential impact of low weight and medication status on our results, we compared AN participants with a BMI below 18.5 kg/m^2 (n = 25; 59.5% of sample) to AN participants with a BMI above 18.5 kg/m^2 (n = 17; 41.5% of sample), and AN participants on medication (n = 25; 61% of sample) to AN participants not on medication (n = 16; 39% of sample) on clinical measures using Welch's two sample t-tests and repeated the rmANOVAs described above for each subsample. Small samples precluded analysis of medication class (Table 1).

RESULTS

Sample Characteristics

AN and HC groups did not differ in age or education (Table 1). AN had significantly lower current BMI ($p < .001$). In AN, there was a significant increase in BMI from treatment admission to discharge (t(39) = 7.9, $p < .001$, Cohen's d = 1.0).

Behavioral Performance

A Group × Block × Set rmANOVA on optimal responses for reward trials revealed a main effect of Block, indicating increased accuracy over time across all participants, consistent with learning, $F(3,225) = 41.482$, $p < .001$, $\eta^2_p = .356$ (Figure 3A). We detected a Group × Block interaction, corresponding to faster learning rates in the HC group compared to AN, $F(3,225) = 5.771$, $p = .001$, $\eta^2_p = .071$. A Group × Set interaction indicated AN were more accurate than HC on Set 1, but less accurate than HC on Set 2, $F(1,75) = 5.556$, $p = .021$, $\eta^2_p = .069$.

For punishment trials, a Group × Block × Set rmANOVA revealed a main effect of Block, indicating increased accuracy over time, $F(3,225) = 3.711$, $p = .012$, $\eta^2_p = .047$ (Figure 3B). A main effect of Group indicated AN performed worse than HC, $F(1,75) = 6.833$, $p = .011$, $\eta^2_p = .083$. Taken together, both groups demonstrated greater accuracy over time (aka, learning) for reward and punishment trials; compared to HC, AN had slower overall learning on reward trials, with better overall accuracy on Set 1 and worse accuracy on Set 2 (possibly suggesting greater difficulty set-shifting and learning new associations, see (Filoteo et al., 2014)), and were less accurate across punishment trials.

Model Generated Parameters

Prediction error learning rates (η)—A Group × Set × PE learning rate type (η_p vs. η_n) rmANOVA revealed a main effect of Group, indicating that AN learned more slowly than HC following both positive PEs and negative PEs, $F(1,75) = 5.521$, $p = .021$, $\eta^2_p = .061$ (Table 3; Figure 4A). A main effect of PE type revealed faster learning rates following positive PEs compared to negative PEs across the entire sample, $F(1,75) = 78.792$, $p < .001$, $\eta^2_p = .512$. That is, faster learning occurred when the outcomes were better than expected relative to when the outcomes were worse than expected.

Prediction error size—To directly examine whether groups might have differed in accuracy as a result of better than or worse than expected outcomes on reward and punishment trials. Group × Set × PE type rmANOVAs for average PE size revealed no effects involving Group for reward trials (all $\eta^2_p < .025$) or for punishment trials (all $\eta^2_p < .045$).

Explore-exploit strategy (β)—A Group × Set rmANOVA for the explore-exploit parameter, β, revealed a main effect of Group, whereby AN had smaller β values than HC, $F(1,75) = 6.366$, $p = .014$, $\eta^2_p = .078$ (Table 3; Figure 4B). Since smaller values imply individuals are exploring more than exploiting stimulus-response-outcome hypotheses, results indicate AN may less decisively make choices.

Choice bias parameters—To assess whether groups differed in the degree to which early reward and punishment reinforcement trials reflected choice biases, the Group × Set interaction for bias values was significant only for reward trials, indicating that HC had a greater bias against making the optimal choice on Set 1, whereas AN had a greater bias against making the optimal choice on Set 2, $F(1,75) = 10.651$, $p = .002$, $\eta^2_p = .124$ (Table 3; Figure S10). This is consistent with the behavioral response data indicating that AN

outperformed HC on Set 1 and performed worse than HC on Set 2 on reward trials. No significant effects of choice bias were detected for punishment trials (all $\eta^2_p < .018$).

Exploratory Clinical Associations

No associations between reinforcement learning model parameters and clinical variables were detected in AN (uncorrected $p < .05$). Separate hierarchical linear regression models indicated the size of positive PE and of negative PE on punishment trials in Set 1 significantly added to the prediction of discharge BMI controlling for admission BMI, treatment length, and medication status (positive PE: multiple $R^2 = .62$, F_change(1,34) = 9.528, $p = .004$; negative PE: multiple $R^2 = .56$, F_change(1,34) = 15.901, $p < .001$). Both models remained significant after Bonferroni correction.

To test whether both positive and negative PE predicted a portion of the change in BMI with treatment, we entered both into the regression model (multiple $R^2 = .64$, F_change(2,33) = 8.546, $p = .001$). Negative PE (Beta = −.348, t = −2.475, $p = .019$) more potently predicted discharge BMI than did positive PE (Beta = −.141, t = −1.063, $p = .296$) (Figure 4C). In other words, AN with smaller negative PE on punishment trials on Set 1, i.e., values closer to −1.0, gained the most weight. Negative PE will approach −1 on punishment trials when successful performers learn to expect outcomes that are close to the favorable outcome, coded 0, but instead receive an unfavorable outcome, coded −1. The eight AN participants with negative PE between −.85 and −1.0 in fact had an average expectancy of 0.013 on punishment trials when negative PE occurred (range for entire sample: −.467 to .545) (see Supplement). Moreover, on punishment trials where negative PE occurred, the regression of expectancy values onto negative PE produced a significant negative regression weight (b = −.419, $p = .048$), implying that AN participants with larger negative PE (i.e. closer to zero) had more negative expectancies about avoiding loss.

Sensitivity Analyses

As expected, the low weight group had lower BMI at admission, time of study, and discharge (all ps < .001, all Cohen's ds > 1.0), and showed greater change in BMI during treatment ($p = .01$, Cohen's d = 1.1), but weight status groups did not differ on any other clinical measure. Medication status groups did not differ on any clinical measure, including BMI, change in BMI during treatment, length of treatment, or self-report questionnaires. The rmANOVA results from the full sample reported above were observed in the subsample contrasts. Regression results (PE on punishment trials predicting discharge BMI) were observed only in the low weight sample. Overall, sensitivity analyses suggest weight and medication status did not appreciably contribute to the full sample results.

DISCUSSION

This is the first study to apply computational models of reinforcement learning to assess learning from both reward and punishment in restricting-type AN using an instrumental probabilistic associative learning task. A unique aspect to this study is that we distinctly examined differences in instrumental reinforcement learning from better or worse than expected outcomes by deriving trial-specific PE estimates for both reward and punishment

conditions. We then modeled and compared learning based on positive and negative PEs separately for reward and punishment trials to examine learning rate when a positive PE occurs (unexpectedly favorable outcome) and when a negative PE occurs (unexpectedly disappointing outcome). Model-based results indicated that both HC and AN learn better following positive PEs compared to negative PEs. *Consistent with our hypotheses, individuals with AN have lower learning rates for positive and negative PEs compared to HC.* This indicates that AN learn less than HCs from the same PE, slowing their learning of favorable choices. This deficit in learning to predict the most favorable choice was also evidenced in their optimal choice performance by a flatter learning curve on reward trials and by fewer optimal responses on punishment trials. These results are consistent with previous work showing poorer learning performance from reward-based feedback in ill AN (Foerde & Steinglass, 2017) and extends these findings to learning from loss-based feedback. Deficits in learning from punishment could help explain the rigid persistence of disordered eating behaviors despite negative consequences.

The degree to which cognitive inflexibility and difficulty set-shifting in AN contribute to altered reinforcement learning remains to be determined; assessing reversal learning may inform this issue. The lower explore-exploit β values observed in the AN group suggest that poor learning was not due to perseverative responding, as lower β values indicate that individuals with AN were less decisive about exploiting what they had learned and continued to explore stimulus-response outcomes rather than employing the same strategy across all trials, regardless of whether they were aware of the strategy employed. Clinically, AN is characterized by increased sensitivity to uncertainty (Kesby, Maguire, Brownlow, & Grisham, 2017). It is possible that diminished certainty in exploiting what they learned is secondary to uncertainty in the task contingencies, although this was not directly tested.

In addition to comparing groups on response accuracy and rate of learning, we also examined the size of PE as a determinant of learning level. Counter to our hypotheses, no group differences in magnitude of positive and negative PEs within reward or punishment trials were detected. However, within the AN group, the magnitude of negative PE when punishment was possible was most strongly associated with treatment outcome. Moreover, larger negative PEs were associated with more negative expectations on punishment trials, suggesting that AN individuals who gained the least amount of weight during the course of treatment held negative expectancies about avoiding loss on punishment trials. This negative expectancy is consistent with reports of elevated punishment sensitivity, increased lose-shift behavior on a reversal learning task (Geisler et al., 2017), negative interpretation bias for ambiguous social stimuli that involve the risk of rejection, and tendency to resolve ambiguity in a negative manner in AN (Cardi, Di Matteo, Gilbert, & Treasure, 2014; Cardi, Di Matteo, Corfield, & Treasure, 2012; Cardi et al., 2017). No other learning parameter or clinical measure predicted BMI change during treatment, and PEs were not associated with self-report measures of sensitivity to reward or punishment, suggesting that this learning metric may be a particularly sensitive prognostic indicator.

Other studies have observed a relationship between reward PE brain response and weight gain in AN (DeGuzman, Shott, Yang, Riederer, & Frank, 2017; GKW Frank et al., 2018); for example, elevated absolute PE (positive and negative PE combined) response in the

caudate, orbitofrontal cortex and insula has been associated with less weight gain during inpatient treatment. Taken together, our behavioral findings further support the role of altered PE in the pathophysiology of AN, extending prior findings to include operant learning in response to both reward and punishment, and are consistent with the hypothesis that a failure to appropriately modify expectancies may contribute to poor outcome.

Strengths of this study include novel aspects and refinements of the reinforcement learning model, that included modeling segregating learning for each of the four stimuli within a set, adding parameters to account for choice biases rapidly acquired on early trials, performing Bayesian estimates of model parameters for each subject, and modeling separate positive and negative PE learning rate parameters. However, reinforcement learning models are inherently limited by the parameters included in the model. While our models demonstrated good fit to the behavioral data, future work may consider testing models with additional parameters, such as a stickiness (or perseveration) parameter (Palminteri, Khamassi, Joffily, & Coricelli, 2015). To increase generalizability, we did not exclude for medication use and co-morbidities. Prior studies in major depressive disorder (MDD) report worse learning to reward (Herzallah et al., 2017), and that SSRI antidepressants impair learning from negative feedback (Herzallah et al., 2013). Notably, 50% of our sample was prescribed antidepressants, and 20% of our sample had a comorbid MDD diagnosis. Although our sensitivity analysis suggests medication status did not contribute to overall results, larger, controlled studies are needed to examine the effects of these clinical variables on reinforcement learning. We also do not have neuropsychological data to characterize the general cognitive function of participants; however, groups did not differ on reaction time on the PALT (see Supplement), suggesting the AN group did not have slowed processing speed indicative of cognitive impairment or medication effects. Thus, it is unlikely that differences in reward/punishment learning in AN are reflective of broader cognitive impairment. Lastly, change in BMI is just one metric of treatment outcome; limited data on cognitive symptoms prevented analysis of other outcome measures.

Conclusions

Results suggest that both AN and HC groups learned better following unexpected favorable outcomes (positive PEs) than unexpected disappointing outcomes, suggesting that maximizing positive PEs may potentiate learning in general. Moreover, individuals with AN demonstrated slower learning from both positive and negative experience compared to HC. Additionally, negative PEs on punishment trials were associated with worse treatment outcome. Treatments that modify negative expectations about avoiding loss, or the perceived value of the outcomes themselves, either with medication or cognitive-behavioral strategies, may be effective in promoting recovery. Overall, findings support the potential of applying computational approaches to reinforcement learning in AN to enhance mechanistic explanations of behavior, identify new neurobehavioral constructs relevant to psychopathology and advance treatment development through target identification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Noriko Coburn, Sarah Kouzi, Danika Peterson, and Emily Romero for assistance with participant screening and data collection. In addition, we thank the individuals who participated in this study for their time.

FINANCIAL SUPPORT

This work was supported in part by grants from the National Institute of Mental Health (R01MH113588 to ABG & CEW, R21MH118409 to CEW). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official view of the NIH.

REFERENCES

American Psychiatric Association (2000). Diagnostic & Statistical Manual of Mental Disorders: DSM:VI-TR (4th ed.). Washington, DC: Association AP, editor.

Beck A, Steer R, & Brown G (1996). Beck Depression Inventory—Second Edition. Manual. San Antonio, TX: The Psychological Corporation.

Bernardoni F, Geisler D, King JA, Javadi AH, Ritschel F, Murr J, … Ehrlich S (2018). Altered medial frontal feedback learning signals in anorexia nervosa. Biological Psychiatry, 83(3), 235–243. doi: 10.1016/j.biopsych.2017.07.024. [PubMed: 29025688]

Beylergil SB, Beck A, Deserno L, Lorenz R, Rapp M, Schlagenhauf F, … Obermayer K (2017). Dorsolateral prefrontal cortex contributes to the impaired behavioral adaptation in alcohol dependence. Neuroimage Clinical, 15, 80–94. doi: 10.1016/j.nicl.2017.04.010. [PubMed: 28491495]

Bischoff-Grethe A, McCurdy D, Grenesko-Stevens E, Irvine L, Wagner A, Yau W-Y, … Kaye W (2013). Altered brain response to reward and punishment in adolescents with anorexia nervosa. Psychiatry Research Neuroimaging, 214(3), 331–340. doi: 10.1016/j.pscychresns.2013.07.004.

Bodi N, Keri S, Nagy H, Moustafa A, Myers CE, Daw N, … Gluck M (2009). Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain, 132(Pt 9), 2385–2395. doi: 10.1093/brain/awp094. [PubMed: 19416950]

Brooks S, Rask-Andersen M, Benedict C, & Schioth H (2012). A debate on current eating disorder diagnoses in light of neuro-biological findings: is it time for a spectrum model? BMC Psychiatry, 12, 76. doi: 10.1186/1471-244X-12-76. [PubMed: 22770364]

Brown TA, Cusack A, Anderson LK, Trim J, Nakamura T, Trunko ME, & Kaye WH (2018). Efficacy of a partial hospital programme for adults with eating disorders. European Eating Disorder Review, 26(3), 241–252. doi: 10.1002/erv.2589.

Cardi V, Di Matteo R, Gilbert P, & Treasure J (2014). Rank perception and self-evaluation in eating disorders. The International Journal of Eating Disorders, 47(5), 543–552. doi: 10.1002/eat.22261. [PubMed: 24549635]

Cardi V, Di Matteo R, Corfield F, & Treasure J (2012). Social reward and rejection sensitivity in eating disorders: An investigation of attentional bias and early experiences. The World Journal of Biological Psychiatry, 14(3), 622–633. doi: 10.3109/15622975.2012.665479. [PubMed: 22424288]

Cardi V, Turton R, Schifano S, Leppanen J, Hirsch C, & Treasure J (2017). Biased interpretation of ambiguous social scenarios in anorexia nervosa. European Eating Disorders Review, 25(1), 60–64. doi: 10.1002/erv.2493 [PubMed: 27943534]

Carver C & White T (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS Scales. Journal of Personality and Social Psychology, 67, 319–333. doi: 10.1037/0022-3514.67.2.319

Cloninger C, Przybeck T, Svrakic D, & Wetzel R (1994). The Temperament and Character Inventory (TCI): A Guide to Its Development and Use (Vol. 2, Chapter 4, pp. 19–28). St. Louis, MO: Center for Psychobiology of Personality, Washington University, ISBN 0-9642917-1-1.

Daw ND (2011). Trial-by-trial data analysis using computational models. In Delgado MR, Phelps EA, & Robins TW (Eds.), Decision making, affect, and learning, attention and

performance, Vol. XXIII, (pp. 5–38). Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780199600434.003.0001

DeGuzman M, Shott M, Yang T, Riederer J, & Frank G (2017). Association of elevated reward prediction error response with weight gain in adolescent anorexia nervosa. American Journal of Psychiatry, 174(6), 557–565. doi: 10.1176/appi.ajp.2016.16060671. [PubMed: 28231717]

Fairburn CG & Beglin S (1994). Assessment of eating disorders: interview or self-report questionnaire? The International Journal of Eating Disorders, 16, 363–370. doi: 10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-#. [PubMed: 7866415]

Farrell S & Lewandowsky S (2018). Computational Modeling of Cognition and Behavior. New York: Cambridge University Press.

Filoteo J, Paul E, Ashby F, Frank G, Helie S, Rockwell R,… Kaye W (2014). Simulating category learning and set shifting deficits in patients weight-restored from anorexia nervosa. Neuropsychology, 28(5), 741–751. doi: 10.1037/neu0000055. [PubMed: 24799291]

First M, Spitzer R, Gibbon M, & Williams J (2002). Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P). New York: Biometrics Research, New York State Psychiatric Institute.

Fladung A, Schulze U, Scholl F, Bauer K, & Gron G (2013). Role of the ventral striatum in developing anorexia nervosa. Translational Psychiatry, 3, e315 doi: 10.1038/tp.2013.88. [PubMed: 24150224]

Foerde K & Steinglass J (2017). Decreased feedback learning in anorexia nervosa persists after weight restoration. The International Journal of Eating Disorders, 50(4), 415–423. doi: 10.1002/eat.22709. [PubMed: 28393399]

Frank G, Collier S, Shott M, & O'Reilly R (2016). Prediction error and somatosensory insula activation in women recovered from anorexia nervosa. The Journal of Psychiatry & Neuroscience, 41(2), 304–311. doi: 10.1503/jpn.l50103. [PubMed: 26836623]

Frank G, DeGuzman M, Shott M, Laudenslager M, Rossi B, & Pryor T (2018). Association of brain reward learning response with harm avoidance, weight gain, and hypothalamic effective connectivity in adolescent anorexia nervosa. JAMA Psychiatry, 75(10), 1071–1080. doi: 10.1001/jamapsychiatry.2018.2151. [PubMed: 30027213]

Frank G, Reynolds J, Shott M, Jappe L, Yang T, Tregellas J, & O'Reilly R (2012). Anorexia nervosa and obesity are associated with opposite brain reward response. Neuropsychopharmacology, 37(9), 2031–2046. doi: 10.1038/npp.2012.51. [PubMed: 22549118]

Fritsche M, Mostert P, & de Lange F (2017). Opposite effects of recent history on perception and decision. Current Biology, 27(4), 590–595. doi: 10.1016/j.cub.2017.01.006. [PubMed: 28162897]

Garcia-Perez M & Alcala-Quintana R (2013). Shifts of the psychometric function: Distinguishing bias from perceptual effects. Quarterly Journal of Experimental Psychology, 66(3), 319–337. doi: 10.1080/17470218.2012.708761.

Geisler D, Ritschel F, King J, Bernardoni F, Seidel M, Boehm I, … Ehrlich S (2017). Increased anterior cingulate cortex response precedes behavioural adaptation in anorexia nervosa. Scientific Reports, 7, 42066. doi: 10.1038/srep42066. [PubMed: 28198813]

Gershman S (2016). Empirical priors for reinforcement learning models. Journal of Mathematical Psychology, 71, 1–6. doi: 10.1016/j.jmp.2016.01.006.

Glashouwer K, Bloot L, Veensra E, Franken I, & de Jong P (2014). Heightened sensitivity to punishment and reward in anorexia nervosa. Appetite, 75, 97–102. doi: 10.1016/j.appet.2013.12.019. [PubMed: 24389241]

Glimcher P (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proceedings of the National Academy of Sciences of the United States of America, 108(Suppl 3), 15647–15654. doi: 10.1073/pnas.1014269108. [PubMed: 21389268]

Gold J & Ding L (2013). How mechanisms of perceptual decision-making affect the psychometric function. Progress in Neurobiology, 103, 98–114. doi: 10.1016/j.pneurobio.2012.05.008. [PubMed: 22609483]

Harrison A, O'Brien N, Lopez C, & Treasure J (2010). Sensitivity to reward and punishment in eating disorders. Psychiatry Research, 177(1–2), 1–11. doi: 10.1016/j.psychres.2009.06.010. [PubMed: 20381877]

Harrison A, Treasure J, & Smillie L (2011). Approach and avoidance motivation in eating disorders. Psychiatry Research, 188(3), 396–401. doi: 10.1016/j.psychres.2011.04.022 [PubMed: 21645929]

Hebebrand J & Bulik C (2011). Critical appraisal of the provisional DSM-5 criteria for anorexia nervosa and an alternative proposal. The International Journal of Eating Disorders, 44(8), 665–678. doi: 10.1002/eat.20875. [PubMed: 22072403]

Herzallah M, Khdour H, Taha A, Elmashala A, Mousa H, Taha M, … Gluck M (2017). Depression reduces accuracy while Parkinsonism slows response time for processing positive feedback in patients with Parkinson's Disease with comorbid major depressive disorder tested on a Probabilistic Category-Learning Task. Frontiers in Psychiatry, 8, 84. doi: 10.3389/fpsyt.2017.00084. [PubMed: 28659830]

Herzallah M, Moustafa A, Natsheh J, Abdellatif S, Taha M, Tayem Y, … Gluck M (2013). Learning from negative feedback in patients with major depressive disorder is attenuated by SSRI antidepressants. Frontiers in Integrated Neuroscience, 7, 67. doi: 10.3389/fnint.2013.00067.

Jappe L, Frank G, Shott M, Rollin M, Pryor T, Hagman J, … Davis E (2011). Heightened sensitivity to reward and punishment in anorexia nervosa. The International Journal of Eating Disorders, 44(4), 317–324. doi: 10.1002/eat.20815. [PubMed: 21472750]

Kaye W, Wierenga C, Knatz S, Liang J, Boutelle K, Hill L, & Eisler I (2015). Temperament-based treatment for anorexia nervosa. European Eating Disorders Review, 23(1), 12–18. doi: 10.1002/erv.2330. [PubMed: 25377622]

Keating C, Tilbrook A, Rossell S, Enticott P, & Fitzgerald P (2012). Reward processing in anorexia nervosa. Neuropsychologia, 50(5), 567–575. doi: 10.1016/j.neuropsychologia.2012.01.036. [PubMed: 22349445]

Kesby A, Maguire S, Brownlow R, & Grisham J (2017). Intolerance of uncertainty in eating disorders: An update on the field. Clinical Psychology Review, 56, 94–105. doi: 10.1016/j.cpr.2017.07.002. [PubMed: 28710918]

Linares D, Aguilar-Lleyda D, & Lopez-Moliner J (2019). Decoupling sensory from decisional choice biases in perceptual decision making. eLife, 8, e43994. doi: 10.7554/eLife.43994. [PubMed: 30916643]

Mattfeld A, Gluck M, & Stark C (2011). Functional specialization within the striatum along both the dorsal/ventral and anterior/posterior axes during associative learning via reward and punishment. Learning & Memory, 18(11), 703–711. doi: 10.1101/lm.022889.111. [PubMed: 22021252]

Matton A, Goossens L, Braet C, & Vervaet M (2013). Punishment and reward sensitivity: Are naturally occurring clusters in these traits related to eating and weight problems in adolescents? European Eating Disorders Review, 21, 184–194. doi: 10.1002/erv.2226. [PubMed: 23426856]

Monteleone A, Monteleone P, Esposito F, Prinster A, Volpe U, Cantone E, … Maj M (2017). Altered processing of rewarding and aversive basic taste stimuli in symptomatic women with anorexia nervosa and bulimia nervosa: An fMRI study. Journal of Psychiatric Research, 90, 94–101. doi: 10.1016/j.jpsychires.2017.02.013. [PubMed: 28249187]

Morgan M, Dillenburger B, Raphael S, & Solomon J (2012). Observers can voluntarily shift their psychometric functions without losing sensitivity. Attention, Perception & Psychophysics, 74(1), 185–193. doi: 10.3758/sl3414-011-0222-7.

Moustafa A, Gluck M, Herzallah M, & Myers C (2015). The influence of trial order on learning from reward vs. punishment in a probabilistic categorization task: experimental and computational analyses. Frontiers in Behavioral Neuroscience, 9, 153. doi: 10.3389/fnbeh.2015.00153. [PubMed: 26257616]

Myers C, Moustafa A, Sheynin J, Vanmeenen K, Gilbertson M, Orr S, … Servatius R (2013). Learning to obtain reward, but not avoid punishment, is affected by presence of PTSD symptoms in male veterans: empirical data and computational model. PLoS One, 8(8), e72508. doi: 10.1371/joumal.pone.0072508. [PubMed: 24015254]

National Institute of Mental Health (2016). Behavioral assessment methods for RDoC constructs [Internet], [cited 2020 Nov 28].

Oberndorfer T, Frank G, Fudge J, Simmons A, Paulus M, Wagner A, … Kaye W (2013). Altered insula response to sweet taste processing after recovery from anorexia and bulimia nervosa. American Journal of Psychiatry, 214(2), 132–141. doi: 10.1176/appi.ajp.2013.11111745.

O'Hara C, Schmidt U, & Campbell I (2015). A reward-centered model of anorexia nervosa: A focussed narrative review of the neurological and psychophysiological literature. Neuroscience and Biobehavioral Reviews, 52, 131–152. doi: 10.1016/j.neubiorev.2015.02.012. [PubMed: 25735957]

Palminteri S, Khamassi M, Joffily M, & Coricelli G (2015). Contextual modulation of value signals in reward and punishment learning. Nature Communications, 6, 8096. doi: 10.1038/ncomms9096.

Papadopoulos F, Ekbom A, Brandt L, & Ekselius L (2009). Excess mortality, causes of death and prognostic factors in anorexia nervosa. The British Journal of Psychiatry, 194(1), 10–17. doi: 10.1192/bjp.bp.108.054742. [PubMed: 19118319]

Pearce J & Hall G (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review, 87, 532–552. doi: 10.1037/0033-295X.113.3.584. [PubMed: 7443916]

Plummer M (2017). JAGS version 4.3.0 User Manual. ~https://web.sgh.waw.pl/~atoroj/ekonometria_bayesowska/jags_user_manual.pdf.

Reilly EE, Rockwell RE, Ramirez AL, Anderson LK, Brown TA, Wierenga CE, & Kaye WH (2020). Naturalistic outcomes for a day-hospital programme in a mixed diagnostic sample of adolescents with eating disorders. European Eating Disorders Review, 28(2), 199–210. doi: 10.1002/erv.2716. [PubMed: 31925866]

Rescorla RA & Wagner AR (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In Black AH & Prokasy WF (Eds.), Classical conditioning II: Current research and theory, (pp. 64–99). New York: Appleton Century Crofts.

Roberts M, Tchanturia K, Stahl D, Southgate L, & Treasure J (2007). A systematic review and meta-analysis of set-shifting ability in eating disorders. Psychological Medicine, 37(8), 1075–1084. doi:10.1017/S0033291707009877. [PubMed: 17261218]

Roberts M, Tchanturia K, & Treasure J (2010). Exploring the neurocognitive signature of poor set-shifting in anorexia and bulimia nervosa. Journal of Psychiatric Research, 44(14), 964–970. doi: 10.1016/j.jpsychires.2010.03.001. [PubMed: 20398910]

Schaefer L & Steinglass J (2021). Reward learning through the lens of RDoC: A review of theory, assessment, and empirical findings in the eating disorders. Current Psychiatry Reports, 23(1), 2. doi: 10.1007/sll920-020-01213-9. [PubMed: 33386514]

Schultz W (2016). Dopamine reward prediction error coding. Dialogues in Clinical Neuroscience, 18(1), 23–32. doi: 10.31887/DCNS.2016.18.1/wschultz. [PubMed: 27069377]

Schultz W, Dayan P, & Montague P (1997). A neural substrate of prediction and reward. Science, 275(5306), 1593–1599. doi: 10.1126/science.275.5306.1593. [PubMed: 9054347]

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, … Dunbar GC (1998). The Mini-international neuropspychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. Journal of Clinical Psychiatry, 59(20), 22–33; quiz 34–57.

Shott M, Filoteo J, Jappe L, Pryor T, Maddox W, Rollin M, … Frank G (2012). Altered implicit category learning in anorexia nervosa. Neuropsychology, 26(2), 191–201. doi: 10.1037/a0026771. [PubMed: 22201300]

Sojitra R, Lerner I, Petok J, & Gluck M (2018). Age affects reinforcement learning through dopamine-based learning imbalance and high decision noise-not through Parkinsonian mechanisms. Neurobiology of Aging, 68, 102–113. doi: 10.1016/j.neurobiolaging.2018.04.006. [PubMed: 29778803]

Spielberger C, Gorsuch R, & Lushene R (1970). STAI Manual for the State Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Steinberg EE, Keiflin R, Boivin J, Witten I, Deisseroth K, & Janak P (2013). A causal link between prediction errors, dopamine neurons and learning. Nature Neuroscience, 16(7), 966–973. doi: 10.1038/nn.3413. [PubMed: 23708143]

Strauss M & Smith G (2009). Construct validity: Advances in theory and methodology. Annual Review of Clinical Psychology, 5, 1–25. doi: 10.1146/annurev.clinpsy.032408.153639.

Sutton R & Barto A (2018). Reinforcement Learning: An Introduction (2nd ed.). Cambridge, MA: The MIT Press.

Tchanturia K, Davies H, Roberts M, Harrison A, Nakazato M, Schmidt U, … Morris R (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. PLoS One, 7(1), e28331. doi: 10.1371/journal.pone.0028331. [PubMed: 22253689]

Torrubia R, Avila C, Molto J, & Caseras X (2001). The sensitivity to punishment and sensitivity to reward questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. Personality and Individual Differences, 31(6), 837–862. doi: 10.1016/S0191-8869(00)00183-5.

Urai A, Braun A, & Donner T (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. Nature Communications, 8, 14637. doi: 10.1038/ncommsl4637.

Wagner A, Aizenstein H, Venkatraman M, Fudge J, May J, Mazurkewicz L, … Kaye WH (2007). Altered reward processing in women recovered from anorexia nervosa. American journal of Psychiatry, 164(12), 1842–1849. doi: 10.1176/appi.ajp.2007.07040575. [PubMed: 18056239]

Wagner A, Barbarich-Marsteller NC, Frank GK, Bailer UF, Wonderlich SA, Crosby RD, … Kaye WH (2006). Personality traits after recovery from eating disorders: do subtypes differ? International Journal of Eating Disorders, 39(4), 276–284. doi: 10.1002/eat.20251. [PubMed: 16528697]

Wierenga C, Ely A, Bischoff-Grethe A, Bailer U, Simmons A, & Kaye W (2014). Are extremes of consumption in eating disorders related to an altered balance between reward and inhibition? Frontiers in Behavioral Neuroscience, 9(8), 410. doi: 10.3389/fnbeh.2014.00410.

Wu M, Brockmeyer T, Hartmann M, Skunde M, Herzog W, & Friederich H (2014). Set-shifting ability across the spectrum of eating disorders and in overweight and obesity: a systematic review and meta-analysis. Psychological Medicine, 44(16), 3365–3385. doi: 10.1017/S0033291714000294. [PubMed: 25066267]

Wu M, Brockmeyer T, Hartmann M, Skunde M, Herzog W, & Friederich H (2016). Reward-related decision making in eating and weight disorders: A systematic review and meta-analysis of the evidence from neuropsychological studies. Neuroscience and Biobehavioral Reviews, 61, 177–196. doi: 10.1016/j.neubiorev.2015.11.017. [PubMed: 26698021]

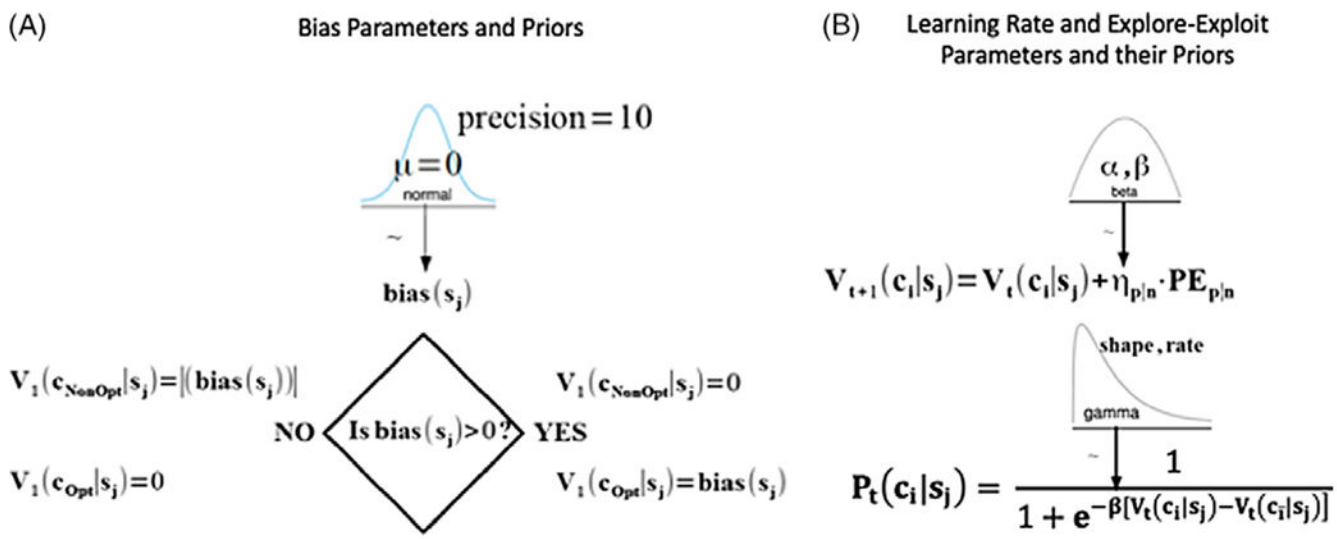

Fig. 1.

(A) Rather than setting all expectancy values, V, to zero on the first trial a stimulus, s_j, is presented, as in the No Bias model, they are set either to a bias value, bias(s_j), or to zero in the First Choice Bias model. The bias(s_j) values are sampled from a normal distribution with mean zero, indicating no bias, and a precision = 10, where precision = 1/variance. If the sampled bias value for stimulus s_j is positive, the choice that would yield the optimal long-term outcome is favored and its expectancy value for trial 1, $V_1(c_{Opt}|s_j)$, is set to the sampled bias value, bias(s_j), whereas the expectancy value for the nonoptimal response, $V_1(c_{NonOpt}|s_j)$, is set to zero. If the sampled bias value is negative the nonoptimal choice is favored and the expectancy value for the nonoptimal choice is set to the absolute value of the bias, whereas the expectancy value for the optimal choice is set to zero. For the First Choice Bias (Singlet) model, the bias parameters for each stimulus is set to the same estimated value bias(s.). (B) The expectancy value for trial $t + 1$ associated with the choice c_i made to stimulus s_j on trial t, $V_{t+1}(c_i|s_j)$, is the expectancy value on trial t updated by the product of a learning rate with the prediction error. Different learning rates, $\eta_{p/n}$, are estimated for positive or negative prediction errors, $PE_{p|n}$. Learning rates are sampled from a beta distribution using values of the α and β parameters listed in Table 2 (Also see Supplement). A logistic equation maps the differences between the expectancy value of the choice made on trial t, $V_t(c_i|s_j)$, and the value of the choice not made, $V_t(\bar{c}_i|S_j)$, to the probability $P_t(c_i|s_j)$ of making the chosen response c_i given that stimulus s_j was presented on trial t. The logistic regression weight β is sampled from a gamma distribution using values of the shape and rate parameters presented in Table 2 (Also see Supplement).

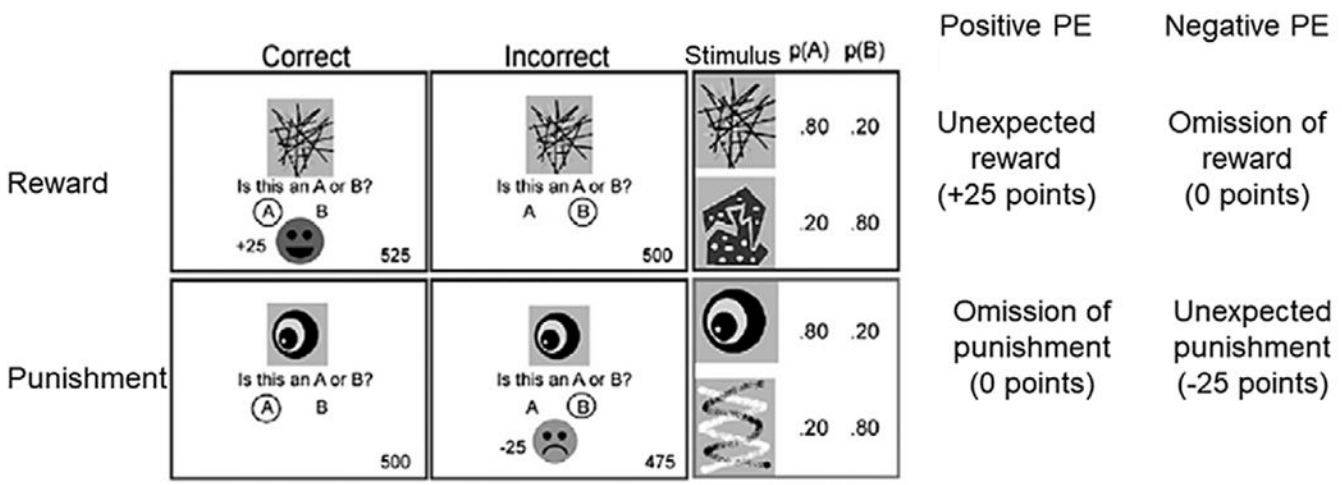

Fig. 2.
Probabilistic associative learning task (copied with permission from (Mattfeld et al., 2011)).

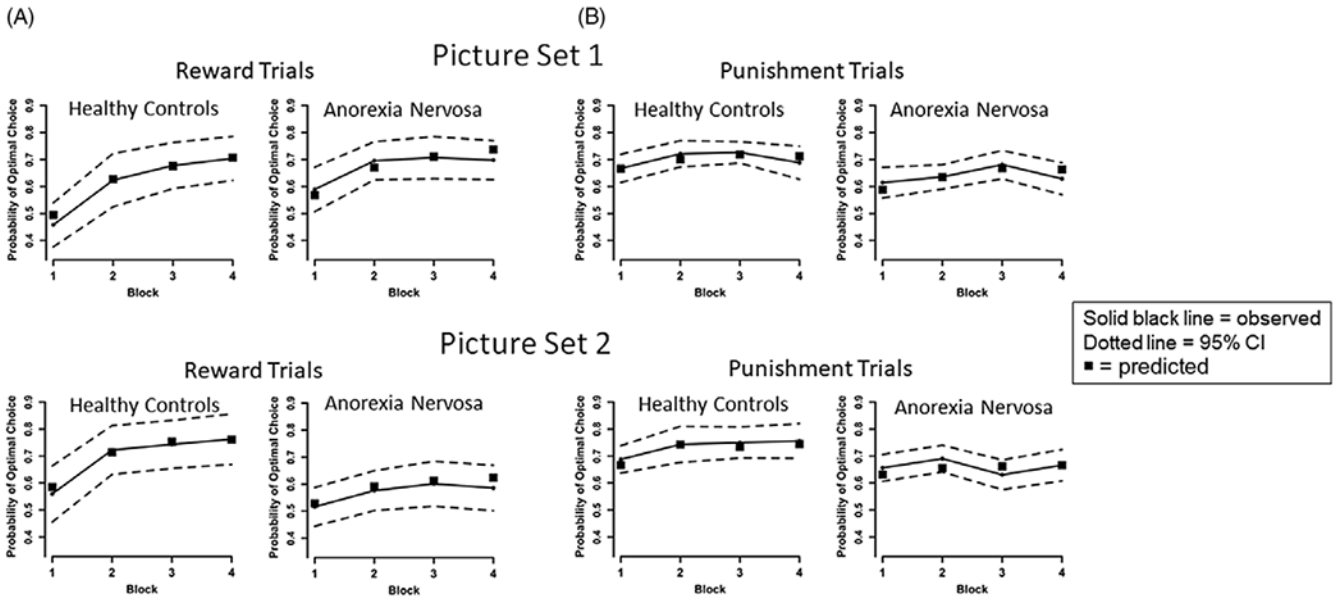

Fig. 3.

Plots of the observed and predicted mean probability of selecting the optimal choice for AN and HC groups across the four blocks by trial type (reward, punishment) and picture set. We calculated for each participant the predicted block means for reward and punishment trials based on the participant's full First Choice Bias model parameter estimates and present the average of these means for AN and HC groups for the two picture sets as black squares. As can be seen, in every instance the model derived means are within the 95% confidence interval of the observed means, and most cover the data means, supporting the prediction model. (A) For observed data, on reward trials, results indicate improved performance over time across all participants, consistent with learning, [main effect of Block, $F(3,225) = 41.482$, $p < .001$, $\eta^2_p = .356$], and the HC group had a greater learning rate overall than the AN group [Group × Block interaction, $F(3,225) = 5.771$, $p = .001$, $\eta^2_p = .071$]. However, AN performed better than HC on Set 1 and worse than HC on Set 2 [Group × Set interaction, $F(1,75) = 5.556$, $p = .021$, $\eta^2_p = .069$]. No other main effects or interactions were significant for reward trials, ps > .3. No other main effects or interactions were significant for reward trials, ps > .3. (B) On punishment trials, performance improved over time across all participants [main effect of Block, $F(3,225) = 3.711$, $p = .012$, $\eta^2_p = .047$], and HC performed better than AN [main effect of Group, $F(1,75) = 6.833$, $p = .011$, $\eta^2_p = .083$]. No other main effects or interactions were significant for punishment trials, ps > .1.

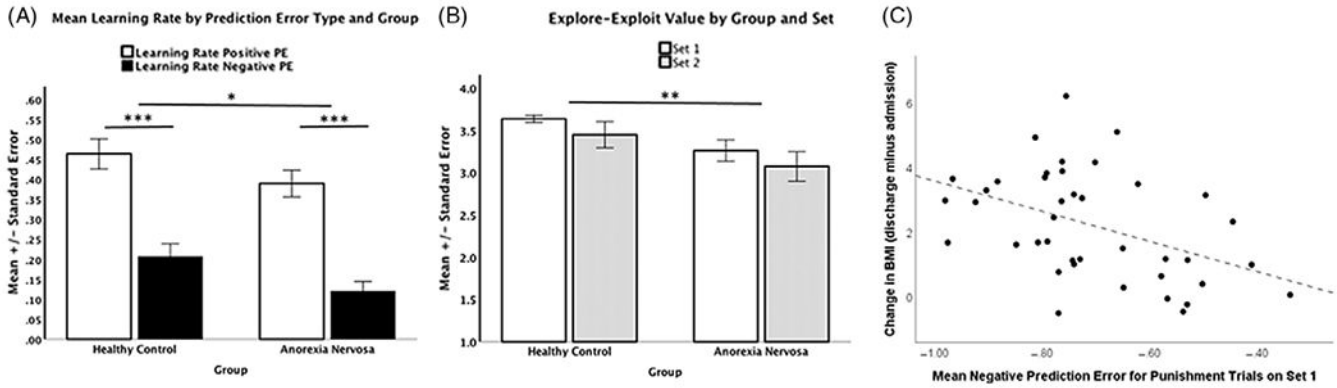

Fig. 4.

(A) Plot of the mean learning rate by prediction error type and group collapsed across set demonstrating the main effect of Group resulting from the Group × Set × PE type ANOVA. The main effect of Group indicated that AN learn more slowly than HC following both positive PEs and negative PEs. A main effect of PE type revealed faster learning rates following positive PEs compared to negative PEs across the entire sample. Neither the main effect of Set nor any interactions were significant (all $\eta^2_p < .039$). (B) Plot of explore-exploit values by group and set showing a main effect of Group. AN had lower β values than HC. Smaller values imply individuals are still exploring stimulus-response-outcome hypotheses and are less certain about exploiting learned rules. The main effect of Set was not significant, nor was the interaction of Group x Set (all $\eta^2_p < .030$). (C) Plot of the change in BMI from admission to discharge with size of negative PE on punishment trials of Set 1. Error bars represent standard error of the mean; $*p < .05$, $**p < .01$, $***p < .001$.

Table 1.

Demographic and clinical characteristics of the sample

	HC (n = 38) Mean (SD)/n (%)	AN-R (n = 42) Mean (SD)/n (%)	t/χ^2 statistic	*p*	Cohen's *d*
Age (years)	21.61 (4.33)	22.81 (9.57)	−.74	.46	.16
BMI at time of study (kg/m^2)	21.65 (2.21)	18.27 (2.19)	6.85	<.001	1.54
Lowest BMI[a]	19.44 (1.64)	15.96 (1.91)	8.56	<.001	1.95
BMI at discharge		20.23 (1.95)			
Education (years)	14.08 (2.72)	13.10 (2.60)	1.65	.10	.37
Female	38 (100)	40 (95)	1.86	.17	
Race			7.07	.07	
Caucasian	26 (68.4)	37 (88.1)			
Asian	9 (23.7)	3 (7.1)			
African American	2 (5.3)	0 (0)			
Other	1 (2.6)	2 (4.8)			
Ethnicity			.04	.85	
Hispanic	6 (15.8)	6 (14.3)			
Non-Hispanic	32 (84.2)	36 (85.7)			
EDE-Q Global Score (n = 40)		3.18 (1.80)			
EDE-Q Restraint		2.64 (1.97)			
EDE-Q Eating Concerns		2.75 (1.72)			
EDE-Q Shape Concerns		3.89 (1.94)			
EDE-Q Weight Concerns		3.52 (2.00)			
STAI State (HC = 35, AN = 39)	25.43 (6.65)	57.25 (13.40)	−13.15	<.001	3.01
STAI Trait (HC = 35)	29.46 (7.31)	56.20 (12.23)	−11.85	<.001	2.65
BDI (n = 40)		27.04 (15.42)			
TCI Harm Avoidance (n = 33)		22.33 (8.49)			
TCI Novelty Seeking (n = 28)		18.07 (8.34)			
BIS Punishment (n = 34)		25.03 (2.75)			
BAS Reward (n = 34)		16.97 (2.15)			
BAS Drive (n = 33)		11.12 (2.13)			
BAS Fun (n = 32)		10.63 (3.05)			

	HC (n = 38) Mean (SD)/n (%)	AN-R (n = 42) Mean (SD)/n (%)	t/χ^2 statistic	*p*	Cohen's *d*
SPSRQ Punishment (n = 17)		21.65 (10.11)			
SPSRQ Reward (n = 17)		21.11 (15.00)			
Comorbid Diagnoses[b]					
Major Depressive Disorder	0	9 (21.4)			
Generalized Anxiety Disorder	0	6 (14.3)			
Panic Disorder	0	0 (0)			
Social Phobia	0	4 (9.5)			
Obsessive-Compulsive Disorder	0	2 (4.8)			
Post-traumatic Stress Disorder	0	6 (14.3)			
Substance Use Disorder	0	2 (4.8)			
Medication class[b,c]					
Antidepressant	0	21 (50)			
Atypical antipsychotic	0	7 (16.7)			
Mood stabilizer	0	3 (7.1)			
Anxiolytic	0	4 (9.5)			
Length of treatment at UCSD (days)		102.7 (46.8)			

Note: Welch's two sample t-tests were used to assess statistical significance for between-group differences in continuous variables. Cronbach's alphas for all self-report measures were strong (α = .84–.99). Self-report questionnaires were completed within 16.1 days of the PALT administration.

[a] Two AN did not complete this assessment.

[b] One AN did not complete this assessment.

[c] Seventeen AN were prescribed only one class of medication, 6 AN were prescribed two classes, and 2 AN were prescribed 3 classes of medication. All medications with presumed dopaminergic action fell within the atypical antipsychotic classification.

BDI = Beck Depression Inventory-Second Edition (BDI-2) (Beck, Steer, & Brown, 1996); BIS/BAS = Behavioral Inhibition/Behavioral Activation Scale (Carver & White, 1994); BMI = body mass index; EDE-Q = Eating Disorder Exam – Questionnaire (Fairburn & Beglin, 1994); SPSRQ = Sensitivity to Punishment Sensitivity to Reward Questionnaire (Torrubia, Avila, Molto, & Caseras, 2001); STAI = Spielberger State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970); TCI = Temperament and Character Inventory (TCI; (Cloninger, Przybeck, Svrakic, & Wetzel, 1994).

Table 2.

Parameters estimated for each of the four models and their prior distributions

Model	Explore-Exploit	Learning Rate	Initial Bias
No Bias	~ gamma(124.4920, 35.2834) Mode = 3.50, SD = 0.3162	η_p, η_n ~ beta(1.5,1.5) Mode = 0.50, SD = 0.25	Fixed to 0.0
First Choice Bias	~ gamma(124.4920, 35.2834) Mode = 3.50, SD = 0.3162	η_p, η_n ~ beta(1.5,1.5) Mode = 0.50, SD = 0.25	$Bias_{r1}$, $Bias_{r2}$, $Bias_{p1}$, $Bias_{p2}$ ~ Norm(0, 10) Mode = 0.0, SD = 0.3162
First Choice Bias (Singlet)	~ gamma(124.4920, 35.2834) Mode = 3.50, SD = 0.3162	η_p, η_n ~ beta(1.5,1.5) Mode = 0.50, SD = 0.25	$Bias_{r1}$ = $Bias_{r2}$ = $Bias_{p1}$ = $Bias_{p2}$ ~ Norm(0, 10) Mode = 0.0, SD = 0.3162

Note. Parameters η_P and η_N represent the learning rates for positive and negative prediction errors respectively. Parameter $Bias_{r1}$ is the bias weight for the first reward stimulus; $Bias_{r2}$ is the bias weight for the second reward stimulus; $Bias_{p1}$ the bias weight for the first punishment stimulus; $Bias_{p2}$ the bias weight for the second punishment stimulus. ~ signifies “distributed as.” The Gaussian distribution in rjags is parameterized as mean and precision, where precision = 1/variance.

Table 3.

Reinforcement learning model generated parameters by group and set

	HC (n = 38) Mean (SD)		AN-R (n = 42) Mean (SD)					
	Set 1	*Set 2*	*Set 1*	*Set 2*	**comparison**	**F**	***p***	η^2_p
η_p	.45 (.33)	.46 (.30)	.33 (.26)	.44 (.34)	HC > AN	5.52	.02	.061
η_n	.16 (.20)	.25 (.30)	.12 (.21)	.12 (.21)	$\eta_p > \eta_n$	78.79	<.001	.512
Positive PE reward trials	.32 (.25)	.30 (.22)	.33 (.22)	.34 (.27)	ns			
Negative PE reward trials	−.63 (.26)	−.62 (.25)	−.63 (.23)	−.60 (.29)				
Positive PE punishment trials	.15 (.20)	.14 (.14)	.10 (.12)	.10 (.14)				
Negative PE punishment trials	−.65 (.19)	−.67 (.17)	−.71 (.16)	−.66 (.18)				
β	3.64 (.24)	3.45 (.92)	3.26 (.79)	3.07 (1.12)	HC > AN	6.366	.014	.078
Bias values reward trials	−.13 (.17)	−.04 (.15)	−.04 (.17)	−.12 (.14)	Group × Set	10.651	.002	.124
Bias values punishment trials	.09 (.20)	.07 (.21)	.04 (.21)	.09 (.20)	ns			

Note: PE: predication error; η_P: learning rate for positive PE; η_N: learning rate for negative PE; β: "inverse temperature" parameter representing the balance between exploring new choice rules and exploiting the rules learned. Two HC and one AN did not complete Set 2.