SUPPLEMENTARY INFORMATION

Neurocomputational Mechanisms Underlying Differential Reinforcement Learning From Wins and Losses in Obesity With and Without Binge Eating

Waltmann *et al.*

Supplementary Material

1. Supplementary Methods

1.1. Inclusion criteria

Participants were included in the normal weight group if they had a BMI between 18 and 25 and in the obese groups if they had a BMI between 30 and 45. Participants were admitted in the Binge Eating Disorder (BED) group if they met DSM-V criteria or showed a subclinical phenotype with one of criteria 2 through 4 not met OR criterion 1a or 1b unclear. Diagnoses were ascertained using the Eating Disorder Examination Interview (Hilbert et al., 2004).

1.2. Exclusion criteria

Before their first visit, potential participants were screened via telephone and excluded if they reported being over- or underweight, pregnant or breast-feeding, if they had problems with color vision, if they suffered from or reported first-degree family history of epilepsy or schizophrenia, as well as if they reported suffering from diabetes, thyroid dysfunction, dyslexia, or having used psychoactive drugs in the past 3 months. At their first visit, participants were additionally screened for present and past mental health problems using the German version of the SCID IV (Wittchen, 1997) and excluded if they met criteria for any current or past diagnosis except anxiety disorders, major depressive or dysthymic disorder, social conduct disorder, childhood disorders (e.g. enuresis) and tics.

1.3. Task Details

The modified version the probabilistic reversal learning task used in this study consisted of two blocks of 140 trials, in which participants had to learn, through trial and error, which of two cards had a higher chance of yielding a positive outcome and relearn the outcome contingencies after reversals. One card had an 80% probability of delivering a positive outcome, while the other had a 20% probability. These contingencies reversed five times per block (after the 35th, 55th, 70th, 85th, and 105th trial. The order of conditions was randomized. Participants indicated their choices using a button box in the MRI setting and the "n" and "m" keys on a PC during training. After each choice, feedback was displayed for 0.5 seconds: a picture of a 10-cents coin with a green plus sign for wins, a picture of a 0 cents coin for neutral outcomes, and a picture of a 10-cents coin with a red minus sign for losses. Additionally, 1 out of every 5 trials followed a "probabilistic" pattern, where correct choices produced negative feedback. Trials were separated by a variable inter-trial interval with an average duration of 2.5 seconds, during which participants viewed a fixation cross.

Before the task, participants received detailed instructions and were informed that one stimulus would consistently outperform the other, but positive outcomes were probabilistic. They were also made aware of the possibility of reversals, where the previously advantageous stimulus could become disadvantageous and vice versa. To ensure comprehension, participants completed two sets of 20 training trials without reversals and were asked to explain the task to the experimenter. Additional explanations were provided if needed. The paradigm was implemented using Psychtoolbox version 3.0.13 in Octave version 4.2.2. The probabilistic reversal learning task was presented on a white screen using a projector in the MRI setting. For training purposes outside the MRI environment, a separate monitor was used.

1.4. Analysis

1.4.1. Accounting for task phases: different schemes

The task has different parts with stable or changing outcome probabilities, which we expected – normatively – to affect both accuracy and switching behavior (more switching and less accuracy in volatile phases). We thus decided to include those task dynamics in the models to account for this variance and increase power. However, there is no established standard as to how this should be done. We therefore used model selection to arbitrate between four schemes: (1), one which differentiates between an acquisition phase encompassing the trials before the first reversal (35 per block) and a reversal phase covering the remaining trials; (2), one which differentiates between two stable phases covering the trials before the first and after the last reversal (i.e., the first and last 35 trials) and a volatile phase encompassing the remaining trials, (3) one differentiating between pre-reversal trials, i.e., the trials leading up to a reversal (105 trials per block), and post reversal trials, i.e., the 5 trials directly following each reversal (25 trials per block), and (4) one which does not account for task dynamics at all. Model selection based on BICs favoured the pre- and post-reversal trials for accuracy. We therefore took this scheme forward for the stay-switch and perseveration models.

1.4.2. Cross-sectional and longitudinal effects of BMI

To investigate effects of BMI in OB and BED, we repeated all GLMMs with average BMI and change in BMI as continuous predictors, while controlling for group. An example of a model would thus be: $\mathrm{Y} \sim X_1 * X_2 * \mathrm{averageBMI} * \mathrm{changeBMI} + X_1 * X_2 * \mathrm{group} + (X_1 * X_2 * \mathrm{changeBMI}|\mathrm{Subject{:}Session})$. We excluded two outliers for this analysis (change in BMI >3 SD from mean).

1.4.3. Computational models

As mentioned in the main body of the article, we fitted a total of 15 learning models to the data. Of these, 12 were different reinforcement learning models based on Q-Learning (Watkins & Dayan, 1992). The following description is taken verbatim from our previous publication (Waltmann et al., 2023).

$$Q_{a,t+1} = Q_{a,t} + \alpha\left(r - Q_{a,t}\right)$$

> "Here, $Q_{a,t}$ refers to the expected value of an action a at trial t. It is updated at each trial based on the prediction error, i.e., the difference between the feedback just obtained after performing this action, r, and the previous expected value $Q_{a,t}$, to form the new expected value $Q_{a,t+1}$. The learning rate α determines how much recent feedback is weighted over the integrated feedback from previous trials (i.e., a learning rate of 1 would only take the last trial into account). The unchosen option is not updated, which we refer to as single update (SU):
>
> $$Q_{a_{unchosen},t+1} = Q_{a_{unchosen},t}$$
>
> We use a reinforcement sensitivity parameter ρ to quantify choice stochasticity:
>
> $$Q_{a,t+1} = Q_{a,t} + \alpha\left(\rho r - Q_{a,t}\right)$$
>
> The reinforcement sensitivity ρ does this by determining the maximum difference between expected values: the lower ρ, the smaller the difference between expected values. In this way, the reinforcement sensitivity poses a lower bound to choice stochasticity, because we use a softmax decision rule to translate expected values into choice probabilities:
>
> $$p(a_i) = \frac{exp\left(Q_{a_i}\right)}{\sum_{j=1}^{K} exp\left(Q_{a_j}\right)}$$

According to this rule, larger difference between expected values translate into more deterministic choice probabilities. Choice stochasticity may be differentially sensitive to positive and negative feedback, resulting in asymmetric staying and switching (e.g., with higher positive ρ, a person's tendency to stay after positive feedback would be stronger than their tendency to switch after negative feedback). This can be captured in separate reinforcement sensitivity parameters for trials after receiving positive or negative feedback:

$$Q_{a,t+1} = Q_{a,t} + \alpha(\rho_{positive/negative} * r - Q_{a,t})$$

Likewise, learning might be differentially sensitive to positive and negative feedback, resulting in different degrees of value updating after positive and negative feedback. This can be captured in models with different learning rates for positive and negative feedback.

$$Q_{a,t+1} = Q_{a,t} + \alpha_{positive/negative}(\rho * r - Q_{a,t})$$

Because the task has an anticorrelated structure, such that if the chosen stimulus yields positive feedback, the other would have invariably yielded negative feedback, participants can use this information to simultaneously update the expected values of both the chosen and the unchosen option. This can be captured in a double update (DU) model:

$$Q_{a_{unchosen},t+1} = Q_{a_{unchosen},t} + \alpha\left((-\rho * r) - Q_{a_{unchosen},t}\right)$$

Like the original single update (SU) model, this can be extended with separate learning rates for positive and negative feedback. Finally, it is conceivable that individuals use their knowledge of the task structure and perform double updating but do not update the unchosen option as much as the chosen one. This can be captured using a discount weight κ, which attenuates updating of the unchosen option.

$$Q_{a_{unchosen},t+1} = Q_{a_{unchosen},t} + \kappa\alpha\left((-\rho * r) - Q_{a_{unchosen},t}\right)$$

κ can be added to all DU models, changing only the equations for the unchosen option.

The different combinations of parameters – SU or DU, single or separate learning rates and reinforcement sensitivities for positive and negative feedback – yield a total of 12 models […]."

The 13th and 14th models (call them HMM and HMM2r) were Hidden Markov Models, as described in (Schlagenhauf et al., 2014):

> "The *Hidden Markov Model (HMM)* models the hypothesis that subjects choose their action based on their belief $b(s)$ about the underlying state of the task s. The crux in formulating a model based on this, is to assume that in inferring this belief distribution over states, subjects neglect their policy, i.e. they treat action-reward pairs as simple observations $o_t = \{a_t, r_t\}$. The belief variable then takes on a role similar to the Q value in the standard reinforcement learning models above, in the sense that is not really a hidden variable that is being averaged over for action choice, but rather a deterministic function of the past observations and the parameters.
>
> Let the belief over states at time $t+1$, based on all observations (action-reward pairs) up to time t be $p(s_{t+1}|o^t) = p(s_{t+1}|a^t, r^t)$, where $o^t = \{o_\tau\}_{\tau=1}^t = \{a_\tau, r_\tau\}_{\tau=1}^t$. We write:
>
> $$p(s_{t+1}|o^t) = \int ds_t\, p(s_{t+1}|s_t) \frac{p(o_t|s_t)p(s_t|o^{t-1})}{\int ds'_t\, p(o_t|s_t)p(s_t|o^{t-1})}$$
>
> The probability of an observation is given by its 'compatibility' with the state. A reward tells us with probability c that we chose correctly and that the state is the state corresponding to action a. In model HMM R/P, a punishment tells us with probability d that the state is the one not corresponding to action a.
>
> $$p(o_t | s_t) = \frac{1}{2} + \frac{1}{2} \times \begin{cases} +c & (\text{if} \quad a_t = s_t \quad \text{and } r_t = 1) \\ -c & (\text{if} \quad a_t \neq s_t \quad \text{and } r_t = 1) \\ -d & (\text{if} \quad a_t = s_t \quad \text{and } r_t = -1) \\ +d & (\text{if} \quad a_t \neq s_t \quad \text{and } r_t = -1) \end{cases}$$
>
> where$0 \leq c, d \leq 1$. For clarity, an alternative way of expressing this is:
>
> $$p(a = 1\, \&\, r = +1 \,|\, s = 1) = p(a = 2\, \&\, r = +1 \,|\, s = 2) = 0.5 + 0.5c$$
> $$p(a = 2\, \&\, r = -1 \,|\, s = 1) = p(a = 1\, \&\, r = -1 \,|\, s = 2) = 0.5 + 0.5d$$
>
> Note that in the model HMM (as opposed to [HMM2r]), rewards and punishments are treated as equally informative, i.e. $c = d$. [...]
>
> The probability of staying in a state is γ, and hence:
>
> $$p(s_{t+1}|s_t) = \begin{cases} \gamma & if\ s_{t+1} = s_t \\ 1-\gamma & if\ s_{t+1} \neq s_t \end{cases}$$
>
> Finally, the current belief $b = p(s_t = 1|o^{t-1})$ about the state of the task based on past observations is mapped into action probabilities via a sigmoidal function as in the reinforcement learning models, except that the steepness is fixed to 20:
>
> $$p(a_t = 1 | b) = [1 + \exp(-20(b - 1/2))]^{-1}$$
>
> Note that we do not infer the steepness of the sigmoid β as this trades off with the state estimate."

The final model was an experience-weighted attraction model (EWA) as described in (den Ouden et al., 2013):

"The key feature of this model is learning that is weighted by an experience weight. In this model, perseveration on reversal could occur because of an increasing reluctance to update the value of stimuli/choices every time they are chosen. […] [T]he experience-weighted attraction (EWA) model is described by the following equations:

$$n_{c,t} = n_{c,t-1} \times \rho + 1 \qquad \text{(Equation 1)}$$

$$v_{c,t} = \left(\mathrm{v_{c,t-1}} \times \varphi \times n_{\mathrm{c,t-1}} + \lambda_{\mathrm{t-1}}\right)/\mathrm{n_{c,t}} \qquad \text{(Equation 2)}$$

Here, $n_{s,t}$ is the ''experience weight'' of stimulus s […] on trial t, which is updated on every trial, using the experience decay factor ρ. $v_{c,t}$ is the value of choice *c* on trial t, $\lambda_t \in \{0,1\}$ for the outcome received in response to that choice and φ is the decay factor for the previous payoffs, equivalent to the learning rate in the Rescorla-Wagner model. In particular, note that for ρ = 0, $n_{c,t}$ is everywhere 1, and the model reduces to Rescorla-Wagner. For ρ > 0, the experience weights promote more sluggish updating with time. Note that a rearrangement of the parameters is required to see the equivalence between these equations and Rescorla-Wagner. The Rescorla-Wagner learning rate, usually denoted α, is here equivalent to (1 – φ). Moreover, the softmax inverse temperature β […] is equivalent to the product $\beta\alpha$ in Rescorla-Wagner. This is because the values $v_{c,t}$ learned here are scaled by a constant factor of 1/α relative to those learned by their Rescorla-Wagner equivalents. This rescaling makes the model more numerically stable at small α."

1.4.4. Model-fitting

Parameters were estimated using empirical Bayesian estimation implemented in MATLAB R2023a using the emfit toolbox (Huys et al., 2011, 2012; Huys & Schad, 2015). Again, the following description is taken verbatim from our previous publication (Waltmann et al., 2023):

"For fitting, we inverse-logit-transformed the learning rates and double update weight in in order to constrain them to their natural range (0 and 1). For models with a single reinforcement sensitivity, we used an exponential transform to ascertain that it be positive; for models with separate reinforcement sensitivities for wins and losses, the parameters were left in native space."

Similar transformations were applied to the parameters of the other models. For the Hidden Markov Models, all free parameters were inverse-logit-transformed. For the EWA model, β was exponentiated and φ and ρ were inverse-logit-transformed. The following description of the model fitting process is taken verbatim from our previous publication (Waltmann et al., 2023):

> "In standard maximum likelihood estimation, the quantity to be maximized is $log\big(p(data|\theta)\big)$. By contrast, in maximum aposteriori estimation with empirical priors (EM-MAP), a prior on θ is provided, such that the quantity to be maximized becomes $p(data|\theta) * p(\theta)$. This prior is a multivariate gaussian, inferred empirically from the multivariate distribution of the estimates across subjects. It is used iteratively in an expectation maximization procedure (Huys et al., 2012) to find the maximum posterior likelihood of the data. The first expectation step (E-step) is equivalent to maximum likelihood estimation. The resulting parameters are taken forward to the maximisation step (M-step), where they are used to construct the multivariate prior. The E-step is then repeated, although from the second E-step onwards, the product of likelihood and prior is maximised. E- and M-steps are then alternated until no further improvement in $p(data|\theta) * p(\theta)$ can be achieved. To minimize the risk of local minima, we restarted the optimization 3 times (at each E step) at different random starting points, taking the best iteration forward."

We fit separate parameters for the win and loss conditions to capture condition-specific group differences, and one set of parameters across the two sessions of each condition to increase power (number of trials per condition) and match the behavioral analysis, which included session as a nested random factor.

Despite a multi-layered selection procedure (we took the best fitting model out of 10 repetitions of the entire fitting procedure), we had trouble identifying a stable model. Thus, although the results stayed qualitatively the same, posterior likelihoods and parameters values differed slightly each time the analysis was performed. To maximise reproducibility, we therefore reran the fitting procedure for the best fitting model another 10 times, while repeating the each E-step 40 times (Waltmann et al., 2022). While this did not ultimately produce perfectly stable models, we are confident that the best fitting model from that procedure indeed fits the data well, also considering good recoverability (see below).

1.4.5. fMRI

Scanning sequences. All magnetic resonance imaging was performed using a 3 Tesla Siemens MAGNETOM Skyra_fit scanner. Before functional imaging, we collected a field map to account for field inhomogeneities. We then acquired gradient-echo T2*-weighted echo-planar images optimized for blood oxygenation level dependent (BOLD) contrast, using a routine with multi-band acceleration (factor 3). Sixty slices covering the entire brain in oblique orientation at approximately 20 to 30° to AC-PC were acquired in interleaved order, with 2.5mm thickness, 2.5x2.5mm^2 in plane voxel resolution, TR=2s, TE=22ms, and a flip angle of 80°. After functional scanning, we acquired structural images as part of a multi echo (ME) Magnetization-Prepared 2 Rapid Gradient Echoes (MP2RAGE) sequence, which yields T1, T2* and magnetic susceptibility maps (TR=7000ms, Tes at 1.96ms, 5.83ms, 8.78ms, 11.73ms, 15.18ms, Tis at 945ms and 3770ms, flip1=4°, flip2=6°, matrix=240x256, voxel size 1x1x1mm, slices=192). Because Siemens discontinued this sequence after July 2020, we used a T1-weighted ADNI MPRAGE sequence for participants scanned after this date (TR= 2300.0ms, TE= 2.98 ms, TI=900ms, flip angle = 9°, matrix=240x256, voxel size 1x1x1mm, slices=176).

Preprocessing. We preprocessed the fMRI data using SPM12 in MATLAB 2020b (http://www.fil.ion.ucl.ac.uk/spm/software/spm12). First, we set the origin of the functional and structural images to approximately align with the anterior commissure to facilitate later co-registration and normalization. Next, we performed slice-time correction on the functional images and computed voxel-displacement maps based on the field maps. Then, we realigned and unwarped the images, taking into account motion, distortion, and the interaction between motion and distortion. Subsequently, we spatially normalized the images to MNI (Montreal Neurological Institute) space using the normalization parameters obtained during the segmentation of each participant's anatomical scan. Finally, we applied smoothing to the images using an 8mm full width at half maximum isotropic Gaussian kernel. We conducted individual quality checks of the field-map correction, normalization, and head motion adjustments. There were no exclusions due to artifacts, normalization failures or excessive head motion (maximum mean framewise displacement mm in the X, Y and Z directions in any subject: 0.04mm, 0.41mm, 0.27mm). Before conducting the 1st level statistical analysis, we applied a high-pass filter to the data with a cut-off at 128s.

Parametric modulators. Informed by computational modelling, we constructed parametric modulators as follows: first, we derived trial-by-trial PEs from the fitted computational model for each individual. To differentiate the neural representation of actual and inferred (counterfactual) feedback, we computed both single- and double-update prediction errors (SU-PEs and DU-PEs). For the SU- PEs, we used the single-update (SU) model with separate reinforcement sensitivities and learning rates for positive and negative feedback (corresponding to Eq. 1 and 3, without Eq. 2). We set the positive and negative reinforcement sensitivities to 1 and -1, respectively, to ensure PEs were on the same scale and bounded between +1 and -1. This facilitated separating effects of the learning rate and reinforcement sensitivities and avoided issues with the estimation of the correlation between the BOLD signal and PEs (Katahira & Toyama, 2021). To capture additional counterfactual information contained within PEs from the (winning) double-update (DU) model, we generated trial-by-trial PEs from that model and subtracted the SU-PEs (see Reiter et al., 2017 for a similar approach). These SU- and DU-PEs were included as orthogonalized parametric modulators on the feedback regressor. Second, we calculated individual trial-by-trial choice probabilities based on the fitted parameters of the winning DU model (Daw et al., 2006). These choice probabilities are a function of the relative expected values of the two options and can be read as confidence in the upcoming choice. Third, from the choice probabilities, we constructed a control regressor to reflect trial-by-trial model-fit. Choices predicted with below-chance accuracy (<50%) were coded as 1, and 0 otherwise. This regressor was included to remove variance associated with poor model fit. The choice probabilities and model-fit regressors were included as orthogonalized parametric modulators on the cue regressor.

2. Supplementary Results

2.1. Behaviour: descriptive statistics with results of traditional ANOVA

	NW (N=32) ***M (SD)***	***OB (N=32)*** ***M (SD)***	***BED (N=32)*** ***M (SD)***	***p-Value (ANOVA)***
Accuracy	0.80 (±0.08)	0.77 (±0.09)	0.78 (±0.10)	0.36
Accuracy: Win Condition	0.80 (±0.09)	0.76 (±0.10)	0.79 (±0.09)	0.17
Accuracy: Loss Condition	0.81 (±0.07)	0.79 (±0.09)	0.77 (±0.11)	0.38
Accuracy: Pre-Reversal	0.88 (±0.09)	0.84 (±0.11)	0.86 (±0.12)	0.26
Accuracy: Post-Reversal	0.44 (±0.09)	0.47 (±0.09)	0.43 (±0.08)	0.12
Switching	0.16 (±0.09)	0.21 (±0.11)	0.18 (±0.13)	0.25
Switching: Win Condition	0.15 (±0.09)	0.21 (±0.13)	0.18 (±0.14)	0.18
Switching: Loss Condition	0.17 (±0.10)	0.20 (±0.10)	0.19 (±0.12)	0.42
Switching After Positive Feedback	0.06 (±0.07)	0.10 (±0.11)	0.10 (±0.12)	0.32
Switching After Positive Feed-back: Win Condition	0.06 (±0.08)	0.11 (±0.14)	0.09 (±0.13)	0.24

Switching After Positive Feedback: Loss Condition	0.07 (±0.08)	0.09 (±0.10)	0.10 (±0.12)	0.46
Switching After Negative Feedback	0.35 (±0.14)	0.40 (±0.12)	0.35 (±0.13)	0.23
Switching After Negative Feedback: Win Condition	0.34 (±0.14)	0.39 (±0.14)	0.34 (±0.14)	0.37
Switching After Negative Feedback: Loss Condition	0.35 (±0.14)	0.40 (±0.13)	0.35 (±0.13)	0.19
Switching: Pre-Reversal	0.14 (±0.10)	0.19 (±0.12)	0.17 (±0.14)	0.19
Switching: Post-Reversal	0.27 (±0.10)	0.28 (±0.08)	0.27 (±0.10)	0.91
Perseveration	0.12 (±0.10)	0.11 (±0.07)	0.12 (±0.06)	0.63
Perseveration: Win Condition	0.12 (±0.11)	0.12 (±0.09)	0.12 (±0.08)	0.98
Perseveration: Loss Condition	0.12 (±0.10)	0.08 (±0.06)	0.12 (±0.07)	0.08
Perseveration: Pre-Reversal	0.11 (±0.11)	0.10 (±0.10)	0.11 (±0.09)	0.92
Perseveration: Post-Reversal	0.12 (±0.10)	0.11 (±0.06)	0.13 (±0.07)	0.47

We provide the above for completion. However, note that the above inferential statistics are based on traditional ANOVA, which to our understanding is inferior to the mixed-effects models we report in the main manuscript. It is our understanding that where the raw data is binary, it is best modelled using logistic regressions rather than by using aggregate scores (or proportions) and ANOVAS. Moreover, we have shown, in our previous work, that this approach yields more accurate and reliable estimates (Waltmann et al., 2022).

2.2. Computational model selection and model fit

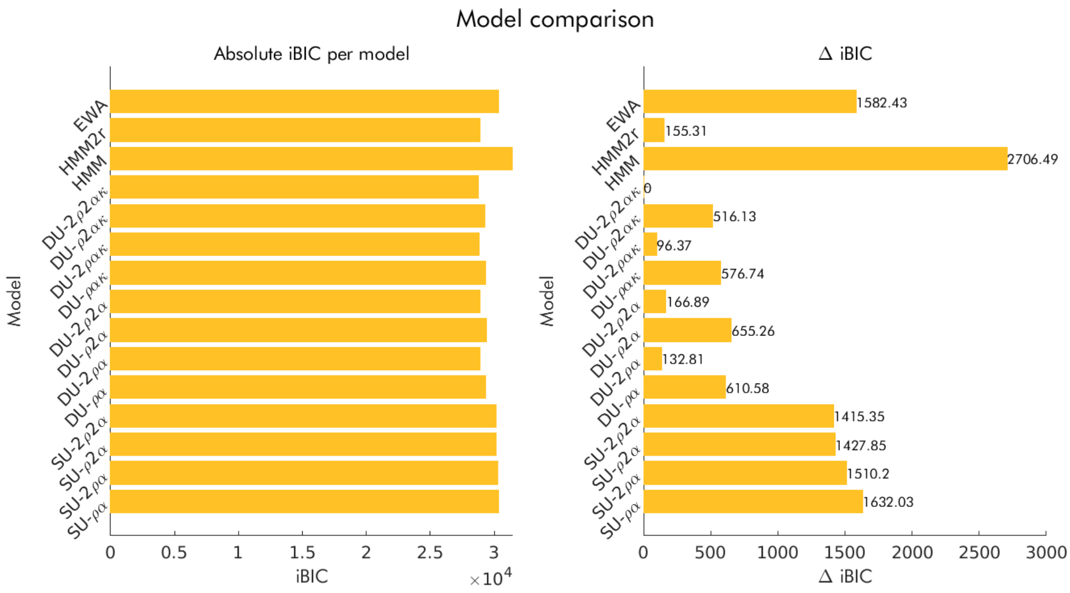

Fig. S1. Model selection. Left panel: absolute iBICs per model. Right panel: difference between the iBIC of the best fitting model and the rest.

The average fit (average trial-by-trial likelihood) of the best fitting model ($\mathrm{DU}2\rho2\alpha\kappa$) was generally good ($\bar{x}$=.74; s=.11). As ascertained by a one-way ANOVA, there were no significant differences in model fit between groups (F(2,93)=1.44, p=.24), Likewise, there were similar numbers of individuals fit at chance per group (OB: 3, BED: 3, NW: 1).

2.3. Recovery

2.3.1. Behaviour.

To ensure good fit at a qualitative level, we used the fitted parameters to simulate 10 new datasets. We then took all simulated trials and submitted them to GLMMs of the form we used in the behavioural analysis, with an additional (nested) random factor for each simulation. We then computed predicted values and correlated them with the predicted values from the original models. Recoverability was generally excellent, with correlation coefficients ranging from r=.83 to r=.98 (Fig. S2). Accuracy in post-reversal trials exhibited slightly worse recoverability with r=.68, perhaps owing to the fact that this metric is very susceptible to noise (Fig. S2).

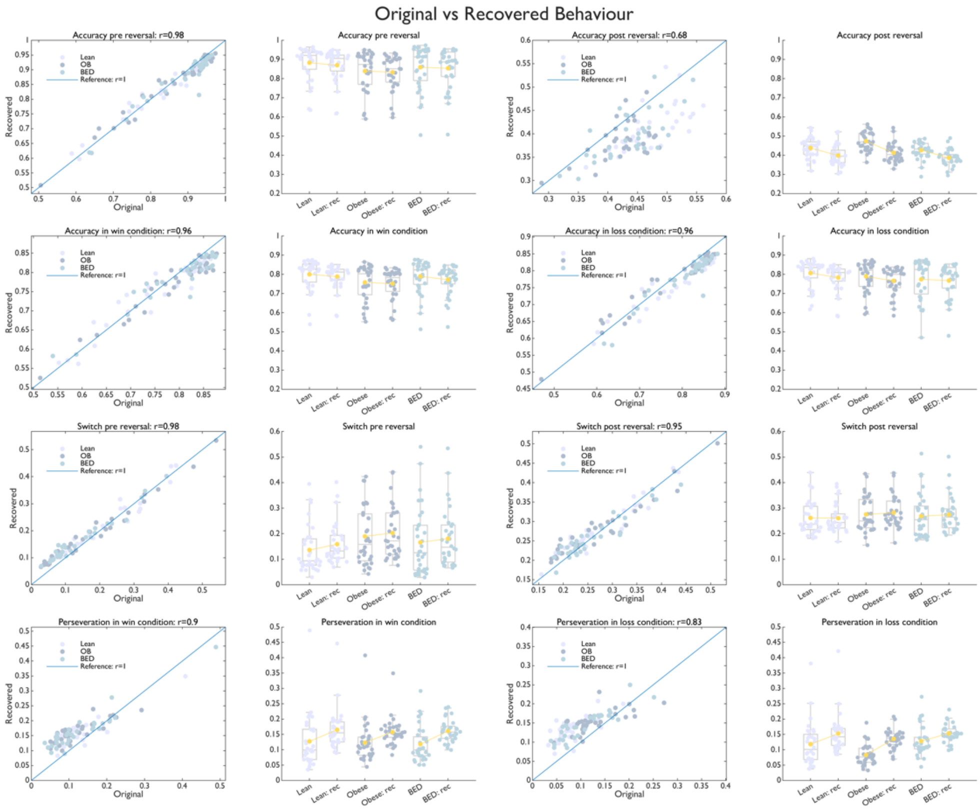

Fig. S2. Recovery analysis. Left to right / top to bottom: Accuracy in pre reversal trials, correlation between original and recovered data. Accuracy in pre reversal trials, boxplots per group for original and recovered data. Accuracy in post reversal trials, correlation between original and recovered data. Accuracy in post reversal trials, boxplots per group for original and recovered data. Accuracy in win condition, correlation between original and recovered data. Accuracy in win condition, boxplots per group for original and recovered data. Accuracy in loss condition, correlation between original and recovered data. Accuracy in loss condition, boxplots per group for original and recovered data. Choice switching in pre reversal trials, correlation between original and recovered data. Choice switching in pre reversal trials, boxplots per group for original and recovered data. Choice switching in post reversal trials, correlation between original and recovered data. Choice switching in post reversal trials, boxplots per group for original and recovered data. Perseveration in win condition, correlation between original and recovered data. Perseveration in win condition, boxplots per group for original and recovered data. Perseveration in loss condition, correlation between original and recovered data. Perseveration in loss condition, boxplots per group for original and recovered data.

2.3.2. Model recovery.

We took the same simulated data forward to perform model recovery. Thus, we fit each of the 15 models described above to each of the 10 simulated datasets. To save computing resources, we repeated the emfit procedure only once per model and simulated dataset (thus slightly diverging from the original fitting procedure as described above). The results show that the correct model was identified as the best fitting model in 8 out of 10 simulated datasets, demonstrating that the model was mostly well recoverable (Fig S3).

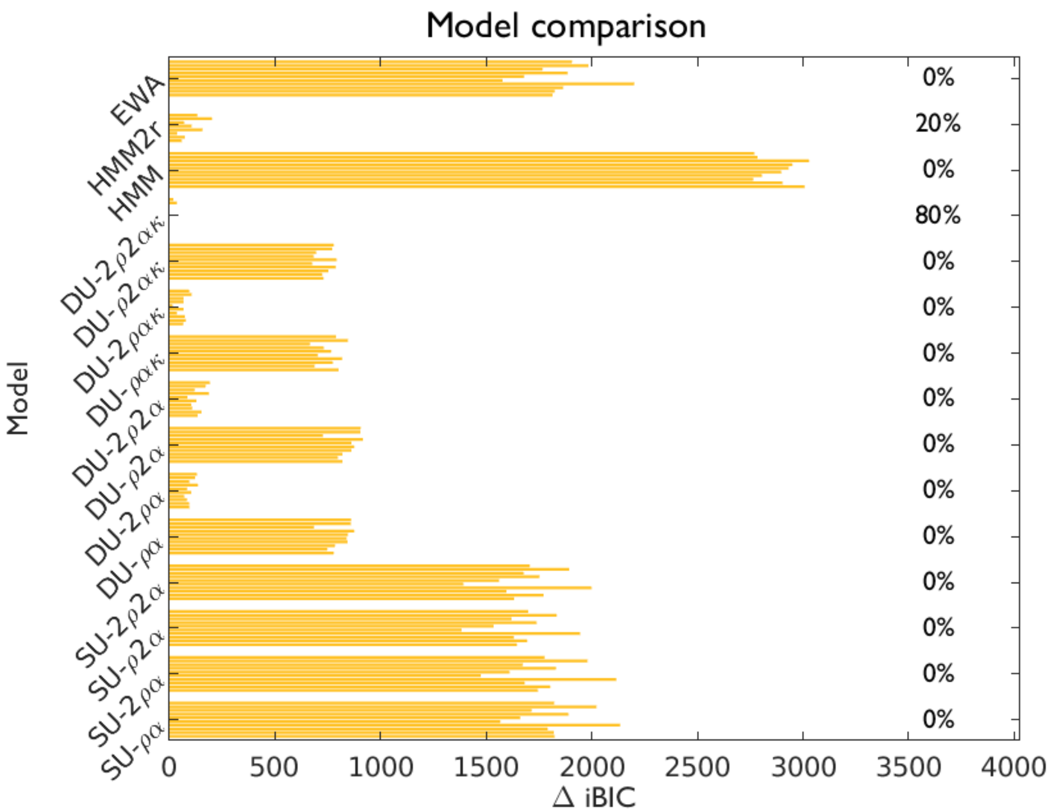

Fig. S3. Model recovery. Ten datasets were simulated on the basis of the fitted parameters of the best fitting model (the weighted double update model with separate learning rates and reinforcement sensitivities for positive and negative feedback (DU2ρ2$\alpha\kappa$)) and all 15 models were fitted to the data. The figure summarises the model selection for all simulations. Each yellow bar represents the difference between the iBIC of each model and the best fitting model, so that the best fitting model has a value of 0. There is one bar per model and simulation. On the right hand side, we plot the percentage of times a model was identified as the best fitting model.

2.3.3. Parameter recovery.

For parameter recovery, we extracted, for each of the 10 simulated datasets, the fitted parameters of the original winning model (the weighted double update model with separate learning rates and reinforcement sensitivities for positive and negative feedback ($\mathrm{DU}2\rho2\alpha\kappa$)). We then averaged the recovered parameters across simulations and correlated them against the original parameters that had been used to generate the data (Fig S4). The parameters generally recovered well, with the exception of the learning rate for positive feedback in the win condition (r=.62), which we believe is still acceptable.

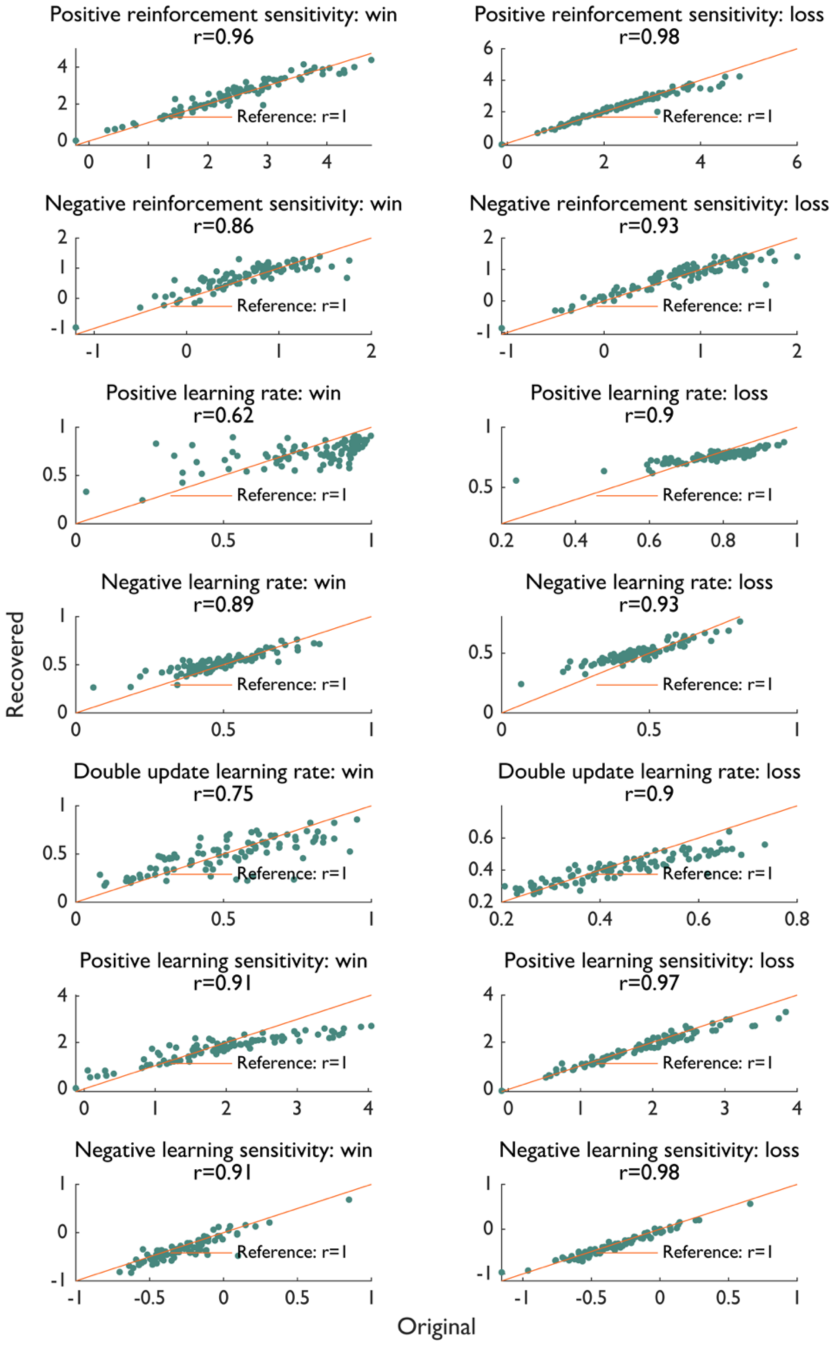

Fig. S4. Parameter recovery. Correlations between the original fitted parameters and the mean parameter values recovered from 10 simulations.

2.4. fMRI

All participants showed BOLD responses reflecting SU-PEs in the striatum and the vmPFC (Fig. S5-A). Specific variance associated with double-update PEs i.e., the variance attributable to counterfactual inference in the double-update model, was associated with BOLD activation mostly in the vmPFC and hippocampus (Fig. S5-B). Trial-by-trial choice probability at cue onset was associated with activity in the vmPFC and frontopolar regions, in addition to the posterior cingulate cortex (Fig. S5-C).

Exploratory analyses revealed no group effects on either choice probabilities or prediction errors correctible at the whole-brain level.

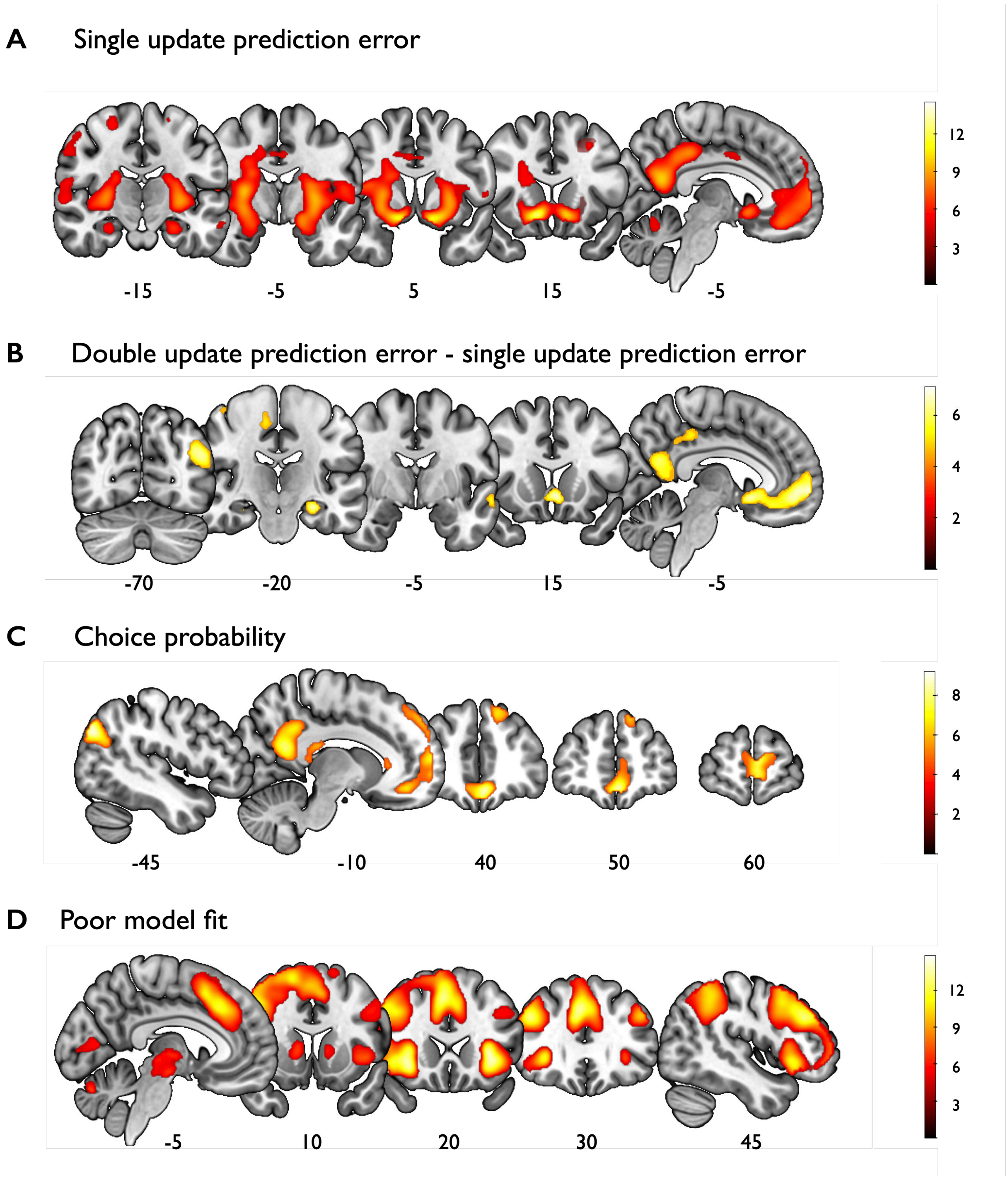

Fig. S5. – A. BOLD response associated with single update prediction errors at feedback onset. B – BOLD response associated with the unique variance associated with prediction errors from the double update model at feedback onset. C– BOLD response associated with choice probability (relative expected values) at trial onset. D – BOLD response associated with poor model fit (choice porbability <.5) at trial onset.

2.4.1. *Positive effect of single update prediction error*

Region	*Side*	*k*	p_{FWE}	*T*	*x,y,z {mm}*
Nucleus accumbens	r	18546	0	14.68	12 10 -10
			0	12.95	-12 8 -12
			0	10.21	-6 -56 18
Inferior frontal gyrus	l	742	0	9.43	-50 32 6
			0	8.03	-34 32 -14
			0	6.93	-22 30 -16
Middle occipital gyrus	l	958	0	8.43	-42 -72 32
Supramarginal gyrus	l	746	0	7.3	-58 -32 30
			0.001	5.41	-52 -30 14
Postcentral gyrus	l	436	0	7.03	-22 -48 66
Inferior frontal gyrus	r	106	0	6.84	52 34 4
Middle temporal gyrus	l	218	0	6.72	-56 -52 -6
Cerebellum	l	239	0	6.24	-26 -44 -28
Lateral front-orbital gyrus	r	39	0	6.01	34 34 -12
Superior frontal gyrus	r	131	0	5.9	24 -12 62
Middle occipital gyrus	r	121	0	5.67	48 -68 26
Middle occipital gyrus	l	62	0	5.62	-28 -92 -4
Cerebellum	l	205	0.002	5.34	-16 -68 -20
			0.003	5.22	-8 -62 -18
			0.008	4.98	0 -52 -34
Precentral gyrus	l	39	0.005	5.1	-22 -20 60
Brain stem	r	6	0.007	5.03	18 -34 -34
Middle temporal gyrus	l	21	0.009	4.98	-62 -14 -18
Precentral gyrus	l	14	0.009	4.96	-58 2 30
Cerebellum	r	12	0.016	4.83	22 -72 -28

Medial frontal gyrus	r	9	0.019	4.79	8 -10 58
Middle temporal gyrus	r	2	0.027	4.69	62 -8 -20

2.4.2. Positive effect of double update prediction error

Region	*Side*	*k*	p_{FWE}	*T*	*x,y,z {mm}*
Medial frontal gyrus	l	1114	0	7.18	-4 60 -2
			0	6.87	-4 14 -8
			0	6.86	-4 52 -4
Medial frontal gyrus	l	591	0	7.01	-8 -58 16
Hippocampus	l	305	0	6.86	-26 -22 -16
			0	5.85	-32 -36 -12
Angular gyrus	l	583	0	6.65	-44 -72 28
			0.019	4.78	-36 -56 20
Middle temporal gyrus	l	73	0	5.79	-62 -6 -12
Cingulate region	l	233	0	5.75	-8 -34 40
			0.048	4.55	-14 -34 48
Precuneus	r	105	0.001	5.42	10 -50 14
			0.018	4.8	18 -44 12
Supplementary motor area	r	62	0.003	5.21	10 -20 50
Middle temporal gyrus	r	35	0.008	5	62 -4 -12
Precentral gyrus	r	68	0.01	4.94	32 -26 70
			0.014	4.86	38 -26 64
			0.026	4.71	46 -18 60
Precentral gyrus	r	16	0.011	4.92	52 0 4
Lateral front-orbital gyrus	l	5	0.012	4.9	-34 32 -14
Superior frontal gyrus	l	47	0.012	4.9	-22 30 42
Postcentral gyrus	r	17	0.019	4.78	16 -40 50
Supramarginal gyrus	l	6	0.02	4.77	-50 -34 26
Hippocampus	r	5	0.022	4.75	28 -18 -18
Paracentral lobule	l	3	0.043	4.57	-18 -40 54

2.4.3. Positive effect of choice probability

Region	*Side*	*k*	p_{FWE}	*T*	*x,y,z {mm}*
Frontal medial orbital gyrus	l	1871	0	9.24	-2 46 -14
			0	8.66	-6 60 10
			0	8.38	-4 60 0
Angular gyrus	l	596	0	8.9	-46 -72 32
Precuneus	l	1543	0	8.6	-8 -54 20
			0	7.56	-4 -50 30
			0	6.55	-20 -46 14
Corpus callosum	-	157	0	6.94	0 20 10
			0.001	5.48	-10 26 6
Superior temporal gyrus	r	132	0	6.35	66 -30 14
			0.002	5.35	56 -28 20
Precentral gyrus	r	110	0.001	5.59	54 -2 8
			0.004	5.2	62 0 10
Hippocampus	l	19	0.001	5.48	-36 -32 -10
Lateral ventricle	r	227	0.001	5.47	20 -38 16
			0.009	5.01	18 -30 20
Parietal superior gyrus	r	80	0.003	5.31	14 -44 64
Inferior frontal gyrus	l	10	0.004	5.2	-54 32 8
Lateral front-orbital gyrus	l	1	0.005	5.17	-34 32 -14
Middle temporal gyrus	l	13	0.006	5.13	-62 -10 -16
Lateral front-orbital gyrus	l	3	0.006	5.13	-30 32 -14
Postcentral gyrus	l	43	0.009	5.02	-20 -40 66
Precentral gyrus	r	20	0.01	4.99	38 -24 66
Precentral gyrus	r	23	0.017	4.86	46 -12 52
Paracentral lobule	l	4	0.025	4.75	-14 -30 72
Precentral gyrus	l	5	0.026	4.74	-58 -4 8
Middle temporal gyrus	r	5	0.034	4.67	64 -6 -14
Caudate nucleus	r	3	0.043	4.61	18 -8 28
Postcentral gyrus	r	1	0.048	4.57	50 -70 28

Precentral gyrus	r	1	0.049	4.57	2 -22 60

2.4.4. Positive effect of poor model fit

Region	*Side*	*k*	p_{FWE}	*T*	*x,y,z {mm}*
Insula	r	10797	0	14.85	32 26 -2
			0	14.15	6 26 40
			0	13.08	44 30 30
Insula	l	1114	0	14.37	-32 22 -2
Angular gyrus	r	3169	0	11.11	50 -44 44
			0	7.2	10 -64 50
Middle frontal gyrus	l	1373	0	10.23	-42 28 32
			0	8.14	-44 2 34
			0	7.66	-36 54 10
Thalamus	r	2036	0	9.6	12 -8 4
			0	7.91	-8 -10 0
			0	7.8	16 4 -2
Angular gyrus	l	2057	0	9.51	-38 -46 44
			0	6.01	-14 -66 54
Cerebellum	l	454	0	9.14	-38 -58 -30
			0	8.01	-8 -76 -26
Middle temporal gyrus	r	312	0	7.61	58 -32 -8
Cuneus	l	363	0	7.2	-12 -74 8
Middle frontal gyrus	l	276	0	7.2	-26 -2 52
Cerebellum	r	84	0	6.84	36 -56 -30
Cuneus	r	212	0.001	5.72	16 -68 8
			0.003	5.31	14 -78 8
			0.003	5.31	22 -60 4
Superior frontal gyrus	l	57	0.001	5.69	-16 10 64
Frontal superior orbital gyrus	l	1	0.033	4.72	-20 46 -12

2.5. Additional analyses

2.5.1. Choice switching: effects of binge eating frequency.

Results showed a three-way interaction of change in binge eating frequency, previous feedback, and trial type on choice switching (beta=-0.72, t(12807)=-3.73, p<.001, Fig. S6, in addition to the respective 2-way interactions (previous feedback x change in binge eating frequency (beta=-0.53, t(12807)=-2.22, p=.03)), trial type and change in binge eating frequency (beta=-0.45, t(12807)=-3.73, p=.02). As Fig. S6 shows, individuals switched less before reversals after receiving negative feedback when they reported higher binge eating frequency relative to baseline.

2.5.2. Sensitivity analyses: BDI and STAI

To ascertain that the results were not confounded by differences in symptoms of depression or trait anxiety, we repeated our analyses with the BDI and STAI, respectively, as covariates.

Accuracy. There were no significant main effects of or interactions with BDI on accuracy (all *p* > .19). The group differences did not substantially change, although the BED-OB x trial-type effect was reduced to a trend (beta=-0.22, t(41973)=-1.83, p=.07). There were no significant main effects of or interactions with STAI on accuracy (all *p* > .25). The group differences did not substantially change.

Switching. There were no significant main effects of or interactions with BDI on switching (all *p* > .25). The group differences did not substantially change. There were no significant main effects of or interactions with STAI on switching (all *p* > .53). The group differences did not substantially change.

Perseveration. There was a significant main effect of BDI on perseveration (beta=-0.22, t(8550)=-2.32, p=.02). The group differences did not substantially change, although the BED-OB x condition effect was reduced to a trend (beta=-0.25, t(8550)=-1.92, p=.06). There was a significant main effect of STAI on perseveration (beta=-0.20, t(8891)=-2.38, p=.02).The group differences did not substantially change.

Learning rates. There were no significant main effects of or interactions with BDI on learning rates (all *p* > .19). The group differences did not substantially change. Likewise, there were no significant main effects of or interactions with STAI on learning rates (all *p* > .19). The group differences did not substantially change.

Learning sensitivity. There were no significant main effects of or interactions with BDI on learning sensitivity (all *p* > .13). The BED-OB x condition interaction effect did not change, however, the BED-OB x Feedback interaction was now significant (beta=0.19, t(360)=2.14, p=.03). There were no significant main effects of or interactions with STAI on learning sensitivity (all *p* > .46). The group differences did not substantially change.

Double update learning rates. There were no significant main effects of or interactions with BDI on learning rates (all *p* > .38). The group differences did not substantially change. Likewise, there were no significant main effects of or interactions with STAI on learning rates (all *p* > .21). The group differences did not substantially change, although the BED-OB x Feedback interaction was reduced to a trend (beta=0.02, t(340)=1.84, p=.07).

2.5.3. Exploratory analyses: BMI

Results showed a complex interaction between average BMI, change in BMI, and condition, such that accuracy in the loss, but not win condition, decreased with increasing BMI in participants with relatively low average BMI, and increased with increasing BMI in participants with relatively high BMI (beta=-.06, t(26494)=-2.03, p=.04, Fig. S7). However, the interaction between average BMI and change in BMI was not significant by itself in either condition (all p>.05).

Further, there was an interaction between change in BMI and trial type on choice switching, which was moderated by previous feedback (beta=-0.15, t(26013)=-2.4, p=.02). As Fig. S8 shows, weight gain was associated with more switching before reversals, especially after positive feedback, and less switching after reversals. This suggests that weight gain comes with increasingly maladaptive choice switching, consistent with our observation of enhanced pre-reversal choice switching in OB.

Moreover, there was an interaction between average BMI and condition on perseveration (beta=0.12, t(5505)=2.67, p=.02): the higher the average BMI, the more individuals perseverated in the win condition (see Fig. S9). There was no significant effect of change in BMI. This resonates with our finding at the group level, where OB showed enhanced perseveration in the win condition.

2.5.4. Exploratory analyses: UPPS

In the introduction, we speculate that negative urgency might be responsible for a putative difference between learning to obtain rewards vs to avoid losses in BED. Unfortunately, the version of the UPPS-scale we used does not differentiate between positive and negative urgency and is therefore not ideally suited to investigate this question. However, given group differences on this scale, we nonetheless repeated our analyses with the UPPS urgency scale replacing the group factor as a predictor. Results for accuracy echoed the BED-OB x condition effects, such that individuals with higher urgency scores performed worse in the loss condition than individuals with lower urgency scores, though specifically in pre-reversal trials (Trial-type x Condition x UPPS-U: beta=-0.04, t(43883)=-2.51, p=.01). This remained true even when we added the factor group back in, suggesting that the effect was not merely driven by the association between BED and urgency. There were no effects of UPPS-U on the other metrics (switching, perseveration, computational model parameters). The UPPS-Perseveration scale, though also different between groups, was unrelated to behaviour and computational modelling parameters. This is mostly supportive of our hypothesis, however, further targeted, experimental research is needed to clarify the specific role of negative urgency in reinforcement learning for obtaining rewards vs avoiding losses.

3. Supplementary figures

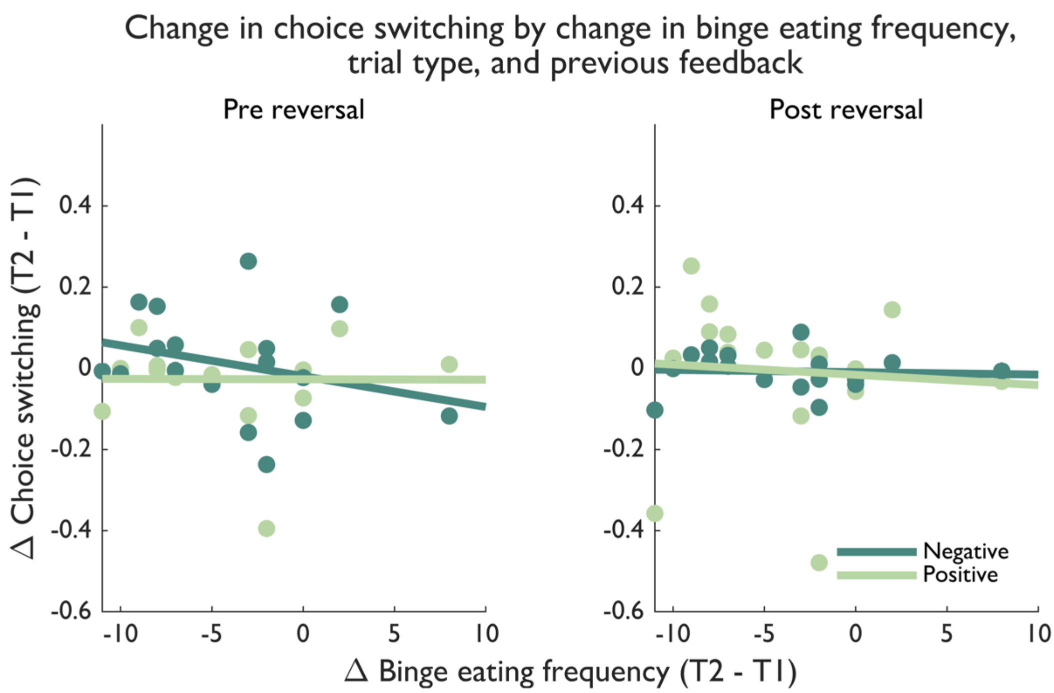

Fig. S6. Change in choice switching by trial type, previous feedback, and change in binge eating frequency between sessions in the BED group. Left panel – Pre reversal trials. Right panel – Post reversal trials. Individual dots represent predicted values from the GLMM.

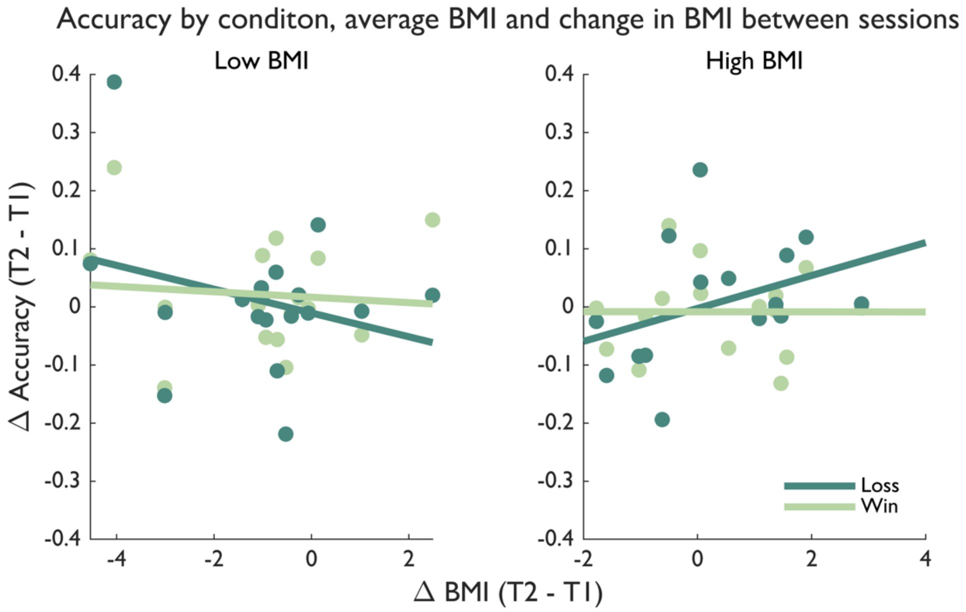

Fig. S7. – Accuracy by condition, average BMI and change in BMI between sessions in the obese groups. Left panel – individuals with relatively low average BMI. Right panel – individuals with relatively high average BMI. Individual dots represent predicted values from the GLMM.

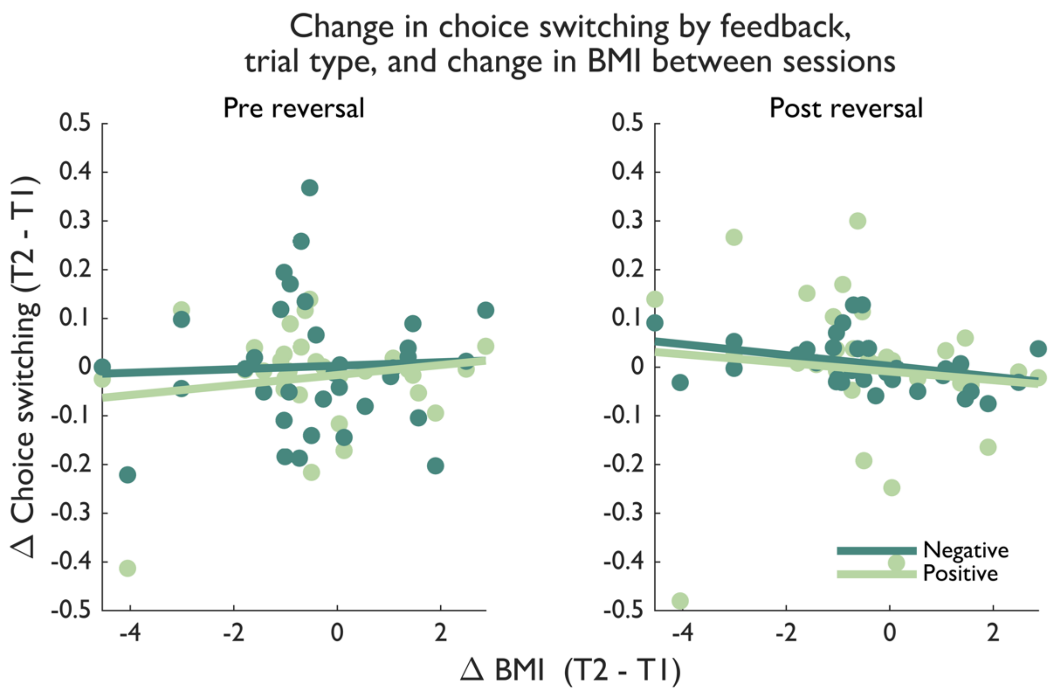

Fig. S8. – Choice switching by trial type, previous feedback and change in BMI between sessions in the obese groups. Left panel – Pre reversal trials. Right panel – Post reversal trials. Individual dots represent predicted values from the GLMM.

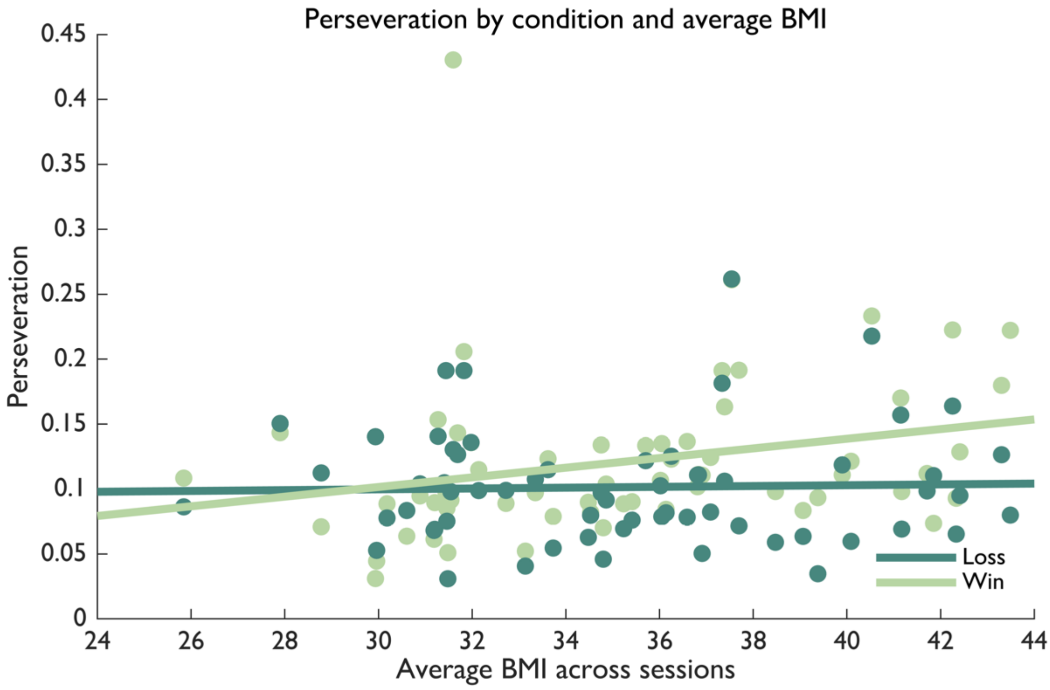

Fig. S9. Perseveration by condition and average BMI across sessions. Individual dots represent predicted values from the GLMM.

4. Supplementary results and figures: sample including subjects without MRI

4.1. Demographics and sample characteristics including subjects without MRI

	NW	*OB*	*BED*	*p-Value*
N	43	43	43	
Age	30.23 (±6.48)	30.54 (±5.77)	30.63 (±7.03)	0.96
BMI	22.55 (±1.91)	35.06 (±3.44)	35.79 (±4.59)	0.41
Follow-up Interval (years)	0.91 (±0.61)	0.95 (±0.46)	0.76 (±0.29)	0.28
Drop-out	23.26 %	30.23 %	30.23 %	0.71
Gender	69.77 %	67.44 %	69.77 %	0.97
Years of education (full-time)	17.13 (±4.45)	17.43 (±5.01)	17.52 (±3.72)	0.91
TMT-A	19.80 (±4.50)	20.25 (±4.98)	21.08 (±6.32)	0.53
TMT-B	41.63 (±10.07)	40.17 (±14.57)	42.39 (±16.54)	0.76
Digit Span Forward	6.67 (±1.30)	6.37 (±1.02)	6.26 (±1.08)	0.23
Digit Span Backwards	5.30 (±1.37)	5.14 (±1.47)	5.02 (±1.07)	0.62
Digit-Symbol-Substitution Task	83.42 (±10.86)	81.77 (±13.69)	75.84 (±14.41)	0.02
Verbal IQ (Wortschatztest)	109.74 (±9.25)	105.30 (±10.06)	104.23 (±6.91)	0.01
Binge episodes (last 28 days)	0.26 (±0.99)	0.37 (±1.62)	6.50 (±5.01)	
EDEQ total	0.75 (±0.86)	1.50 (±1.21)	2.55 (±0.74)	<.001

EDEQ restraint	0.71 (±0.84)	1.34 (±1.20)	1.68 (±1.07)	<.001
EDEQ Eating Concern	0.22 (±0.22)	0.55 (±0.64)	1.91 (±1.01)	<.001
EDEQ Weight Concern	1.82 (±2.07)	3.66 (±2.93)	6.24 (±1.58)	<.001
EDEQ Shape Concern	0.27 (±0.97)	0.43 (±1.19)	0.39 (±0.79)	0.7
BIS 15	29.86 (±6.95)	30.09 (±6.10)	35.76 (±6.80)	<.001
UPPS Urgency	25.36 (±5.43)	26.36 (±5.52)	33.98 (±5.27)	<.001
UPPS Premeditation (-)	21.79 (±4.27)	21.62 (±3.98)	23.26 (±5.19)	0.19
UPPS Perseverance (-)	19.00 (±5.93)	18.55 (±4.28)	22.48 (±4.17)	<.001
UPPS Sensation Seeking	32.38 (±7.33)	31.52 (±7.28)	30.19 (±7.12)	0.38
WBIS	21.88 (±11.22)	36.30 (±13.99)	51.31 (±13.40)	<.001
YFAS	0.17 (±0.44)	0.56 (±1.39)	4.24 (±2.63)	<.001
FCQ	10.55 (±3.78)	11.26 (±3.59)	15.79 (±3.66)	<.001
BDI	3.85 (±4.86)	7.17 (±5.43)	16.79 (±8.35)	<.001
STAI (Trait)	36.56 (±10.41)	37.72 (±9.03)	51.59 (±9.90)	<.001

4.2. Behavioral results including subjects without MRI

4.2.1. Accuracy

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	1.0015	0.0615	0.8809	1.1221	16.2795	0	61370
group_BED	-0.0326	0.087	-0.2031	0.138	-0.3744	0.708	61370
group_Lean	0.0013	0.0862	-0.1676	0.1702	0.0151	0.988	61370
condition	-0.1007	0.0312	-0.1618	-0.0397	-3.2333	0.001	61370
prepost_rev	-1.2044	0.0649	-1.3316	-1.0773	-18.5718	0	61370
group_BED:condition	0.0931	0.0441	0.0067	0.1795	2.1117	0.035	61370
group_Lean:condition	0.041	0.0437	-0.0447	0.1267	0.9376	0.349	61370
group_BED:prepost_rev	-0.0399	0.0917	-0.2197	0.1399	-0.4346	0.664	61370
group_Lean:prepost_rev	-0.08	0.0909	-0.258	0.0981	-0.88	0.379	61370
condition:prepost_rev	0.0005	0.0279	-0.0542	0.0551	0.0175	0.986	61370
group_BED:condition:prepost_rev	-0.0216	0.0394	-0.0989	0.0557	-0.5471	0.584	61370
group_Lean:condition:prepost_rev	0.0371	0.0392	-0.0396	0.1139	0.9486	0.343	61370

4.2.1.1. Effect of binge eating frequency

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	0.9724	0.0727	0.83	1.1148	13.3825	0	18267
condition	-0.0071	0.0313	-0.0684	0.0542	-0.2274	0.82	18267
prepost_rev	-1.2401	0.0755	-1.388	-1.0921	-16.4301	0	18267
crossbingesz	-0.014	0.1228	-0.2546	0.2267	-0.1137	0.91	18267
delta_bez	-0.1444	0.1866	-0.5101	0.2214	-0.7736	0.439	18267
condition:prepost_rev	-0.016	0.0305	-0.0757	0.0437	-0.5244	0.6	18267
condition:crossbingesz	0.0951	0.0537	-0.0101	0.2003	1.7728	0.076	18267
prepost_rev:crossbingesz	0.0861	0.1279	-0.1645	0.3367	0.6734	0.501	18267
condition:delta_bez	-0.0775	0.0788	-0.232	0.0771	-0.9828	0.326	18267

prepost_rev:delta_bez	0.0933	0.1827	-0.2647	0.4514	0.511	0.609	18267
crossbingesz:delta_bez	0.2538	0.384	-0.4989	1.0064	0.6609	0.509	18267
condition:prepost_rev:crossbingesz	-0.0497	0.0509	-0.1495	0.0502	-0.9752	0.33	18267
condition:prepost_rev:delta_bez	0.0178	0.086	-0.1508	0.1865	0.207	0.836	18267
condition:crossbingesz:delta_bez	0.0672	0.1699	-0.2658	0.4001	0.3955	0.692	18267
prepost_rev:crossbingesz:delta_bez	-0.206	0.3753	-0.9417	0.5297	-0.5488	0.583	18267
condition:prepost_rev:crossbingesz:delta_bez	-0.1204	0.1819	-0.4769	0.2361	-0.6619	0.508	18267

4.2.1.2. Effect of BMI

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	0.9372	0.0649	0.8101	1.0644	14.4445	0	37590
group_OB	0.0894	0.0902	-0.0875	0.2663	0.9905	0.322	37590
condition	-0.0288	0.0318	-0.0911	0.0336	-0.9044	0.366	37590
prepost_rev	-1.2019	0.0707	-1.3405	-1.0633	-16.9992	0	37590
crossbmiz	0.0353	0.0461	-0.055	0.1256	0.7663	0.444	37590
delta_bmiz	0.0119	0.0624	-0.1105	0.1343	0.1907	0.849	37590
group_OB:condition	-0.0831	0.0436	-0.1685	0.0023	-1.9072	0.057	37590
group_OB:prepost_rev	-0.0265	0.0982	-0.2189	0.1659	-0.2699	0.787	37590
condition:prepost_rev	-0.0366	0.032	-0.0993	0.026	-1.1466	0.252	37590
condition:crossbmiz	-0.0252	0.0233	-0.0708	0.0204	-1.0835	0.279	37590
prepost_rev:crossbmiz	-0.0093	0.0502	-0.1078	0.0892	-0.1852	0.853	37590
condition:delta_bmiz	-0.0526	0.0329	-0.1171	0.0119	-1.5996	0.11	37590
prepost_rev:delta_bmiz	0.0572	0.0663	-0.0728	0.1871	0.8618	0.389	37590
crossbmiz:delta_bmiz	0.0612	0.0538	-0.0443	0.1667	1.137	0.256	37590
group_OB:condition:prepost_rev	0.0388	0.044	-0.0474	0.125	0.8827	0.377	37590
condition:prepost_rev:crossbmiz	-0.0129	0.023	-0.0579	0.0321	-0.5609	0.575	37590
condition:prepost_rev:delta_bmiz	0.0282	0.0307	-0.0319	0.0883	0.9186	0.358	37590
condition:crossbmiz:delta_bmiz	-0.0747	0.0285	-0.1306	-0.0188	-2.6201	0.009	37590
prepost_rev:crossbmiz:delta_bmiz	-0.0338	0.0574	-0.1462	0.0787	-0.5887	0.556	37590
condition:prepost_rev:crossbmiz:delta_bmiz	0.0012	0.0259	-0.0496	0.0519	0.0446	0.964	37590

4.2.2. Choice repetition

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	-1.9169	0.1154	-2.1431	-1.6907	-16.61	0	60360
group_BED	0.0574	0.1627	-0.2615	0.3763	0.3529	0.724	60360
group_Lean	-0.1062	0.1619	-0.4235	0.2112	-0.6557	0.512	60360
condition	-0.0505	0.0504	-0.1493	0.0483	-1.002	0.316	60360
pfb	-1.3221	0.077	-1.4731	-1.1711	-17.164	0	60360
prepost_rev	0.1495	0.037	0.0771	0.2219	4.0456	<.001	60360
group_BED:condition	0.0333	0.0703	-0.1044	0.171	0.4738	0.636	60360
group_Lean:condition	0.0007	0.0708	-0.1381	0.1394	0.0095	0.993	60360

group_BED:pfb	0.1606	0.1081	-0.0514	0.3725	1.4849	0.138	60360
group_Lean:pfb	0.0249	0.1083	-0.1872	0.2371	0.2305	0.818	60360
condition:pfb	0.0346	0.0405	-0.0448	0.114	0.854	0.393	60360
group_BED:prepost_rev	0.0497	0.051	-0.0503	0.1497	0.974	0.33	60360
group_Lean:prepost_rev	0.0798	0.0517	-0.0216	0.1811	1.5428	0.123	60360
condition:prepost_rev	-0.0513	0.0311	-0.1123	0.0097	-1.6475	0.1	60360
pfb:prepost_rev	-0.008	0.0363	-0.0791	0.063	-0.2215	0.825	60360
group_BED:condition:pfb	-0.0265	0.0559	-0.136	0.083	-0.4747	0.635	60360
group_Lean:condition:pfb	-0.0486	0.0569	-0.16	0.0628	-0.8547	0.393	60360
group_BED:condition:prepost_rev	0.0789	0.0425	-0.0044	0.1622	1.8574	0.063	60360
group_Lean:condition:prepost_rev	0.051	0.0436	-0.0344	0.1364	1.1712	0.242	60360
group_BED:pfb:prepost_rev	0.0109	0.05	-0.0871	0.1089	0.2171	0.828	60360
group_Lean:pfb:prepost_rev	-0.013	0.0507	-0.1124	0.0865	-0.2555	0.798	60360
condition:pfb:prepost_rev	-0.0264	0.0306	-0.0864	0.0336	-0.8625	0.388	60360
group_BED:condition:pfb:prepost_rev	0.0796	0.0417	-0.0021	0.1614	1.9086	0.056	60360
group_Lean:condition:pfb:prepost_rev	0.0525	0.0428	-0.0315	0.1365	1.2253	0.221	60360

4.2.2.1. Effect of binge eating frequency

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	-1.872	0.1363	-2.1392	-1.6047	-13.731	0	17975
condition	-0.014	0.0432	-0.0986	0.0706	-0.3242	0.746	17975
pfb	-1.1907	0.0837	-1.3548	-1.0265	-14.22	0	17975
prepost_rev	0.1361	0.0396	0.0586	0.2137	3.4399	6E-04	17975
crossbingesz	0.1026	0.234	-0.356	0.5612	0.4386	0.661	17975
delta_bez	-0.2756	0.271	-0.8069	0.2556	-1.0171	0.309	17975
condition:pfb	0.0161	0.0389	-0.0603	0.0924	0.4124	0.68	17975
condition:prepost_rev	0.0301	0.0352	-0.0389	0.0992	0.8554	0.392	17975
pfb:prepost_rev	-0.0515	0.0465	-0.1427	0.0396	-1.1084	0.268	17975
condition:crossbingesz	0	0.0749	-0.1467	0.1468	0.0001	1	17975
pfb:crossbingesz	-0.0922	0.1433	-0.3731	0.1887	-0.6435	0.52	17975
prepost_rev:crossbingesz	0.0209	0.0696	-0.1155	0.1574	0.3006	0.764	17975
condition:delta_bez	-0.2552	0.1169	-0.4843	-0.0261	-2.1834	0.029	17975
pfb:delta_bez	-0.1872	0.1682	-0.517	0.1425	-1.113	0.266	17975
prepost_rev:delta_bez	-0.2747	0.1021	-0.4749	-0.0746	-2.6902	0.007	17975
crossbingesz:delta_bez	0.5344	0.5876	-0.6175	1.6862	0.9093	0.363	17975
condition:pfb:prepost_rev	0.0505	0.0329	-0.0141	0.1151	1.5323	0.126	17975
condition:pfb:crossbingesz	0.0238	0.0676	-0.1088	0.1563	0.3519	0.725	17975
condition:prepost_rev:crossbingesz	0.1236	0.0632	-0.0004	0.2475	1.9541	0.051	17975
pfb:prepost_rev:crossbingesz	0.1297	0.0818	-0.0307	0.29	1.5849	0.113	17975
condition:pfb:delta_bez	-0.178	0.1027	-0.3794	0.0234	-1.7321	0.083	17975
condition:prepost_rev:delta_bez	-0.1623	0.0903	-0.3392	0.0146	-1.7981	0.072	17975
pfb:prepost_rev:delta_bez	-0.3135	0.1036	-0.5166	-0.1103	-3.0248	0.003	17975
condition:crossbingesz:delta_bez	-0.1087	0.2759	-0.6494	0.4321	-0.3939	0.694	17975
pfb:crossbingesz:delta_bez	0.2906	0.3678	-0.4304	1.0116	0.7901	0.43	17975

prepost_rev:crossbingesz:delta_bez	0.4255	0.2565	-0.0771	0.9282	1.6593	0.097	17975
condition:pfb:prepost_rev:crossbingesz	0.1248	0.0596	0.0081	0.2416	2.0954	0.036	17975
condition:pfb:prepost_rev:delta_bez	-0.086	0.0845	-0.2516	-0.0797	-1.0173	0.309	17975
condition:pfb:crossbingesz:delta_bez	-0.2164	0.2523	-0.711	0.2782	-0.8575	0.391	17975
condition:prepost_rev:crossbingesz:delta_bez	-0.1659	0.2262	-0.6093	0.2776	-0.7332	0.463	17975
pfb:prepost_rev:crossbingesz:delta_bez	0.6198	0.2506	0.1286	1.1109	2.4732	0.013	17975
condition:pfb:prepost_rev:crossbingesz:delta_bez	-0.1321	0.2161	-0.5556	0.2914	-0.6116	0.541	17975

4.2.2.2. Effect of BMI

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	-1.8767	0.1221	-2.116	-1.6374	-15.373	0	36949
group_OB	-0.0827	0.1693	-0.4146	0.2492	-0.4882	0.625	36949
condition	-0.0536	0.0485	-0.1487	0.0415	-1.1055	0.269	36949
pfb	-1.1996	0.0805	-1.3574	-1.0419	-14.905	0	36949
prepost_rev	0.1558	0.0378	0.0818	0.2298	4.1275	0	36949
crossbmiz	0.0299	0.0883	-0.1432	0.203	0.3386	0.735	36949
delta_bmiz	-0.0421	0.1223	-0.2818	0.1977	-0.3437	0.731	36949
group_OB:condition	0.0127	0.0687	-0.1221	0.1474	0.1846	0.854	36949
group_OB:pfb	-0.1197	0.1114	-0.338	0.0986	-1.0749	0.282	36949
condition:pfb	-0.0111	0.0391	-0.0878	0.0656	-0.2841	0.776	36949
group_OB:prepost_rev	-0.0045	0.0533	-0.1089	0.0999	-0.0848	0.932	36949
condition:prepost_rev	-0.0173	0.0324	-0.0808	0.0462	-0.5335	0.594	36949
pfb:prepost_rev	-0.0266	0.0407	-0.1064	0.0533	-0.6521	0.514	36949
condition:crossbmiz	-0.0314	0.0355	-0.1011	0.0382	-0.8841	0.377	36949
pfb:crossbmiz	0.0405	0.0592	-0.0754	0.1565	0.6854	0.493	36949
prepost_rev:crossbmiz	0.0315	0.0282	-0.0238	0.0868	1.1175	0.264	36949
condition:delta_bmiz	0.0117	0.0529	-0.092	0.1153	0.2204	0.826	36949
pfb:delta_bmiz	-0.0864	0.0862	-0.2555	0.0826	-1.0024	0.316	36949
prepost_rev:delta_bmiz	-0.076	0.0426	-0.1594	0.0075	-1.784	0.074	36949
crossbmiz:delta_bmiz	-0.0387	0.1055	-0.2454	0.1681	-0.3665	0.714	36949

group_OB:condition:pfb	0.0415	0.0561	-0.0685	0.1515	0.7402	0.459	36949
group_OB:condition:prepost_rev	-0.0629	0.0471	-0.1552	0.0294	-1.3362	0.182	36949
group_OB:pfb:prepost_rev	-0.0128	0.0577	-0.126	0.1004	-0.2218	0.825	36949
condition:pfb:prepost_rev	0.0147	0.0309	-0.0459	0.0754	0.4757	0.634	36949
condition:pfb:crossbmiz	-0.0235	0.0287	-0.0798	0.0327	-0.8198	0.412	36949
condition:prepost_rev:crossbmiz	-0.0234	0.0239	-0.0704	0.0235	-0.9793	0.328	36949
pfb:prepost_rev:crossbmiz	0.0094	0.03	-0.0494	0.0683	0.3137	0.754	36949
condition:pfb:delta_bmiz	-0.0003	0.0425	-0.0835	0.083	-0.0063	0.995	36949
condition:prepost_rev:delta_bmiz	-0.062	0.0361	-0.1328	0.0088	-1.7172	0.086	36949
pfb:prepost_rev:delta_bmiz	-0.0538	0.0433	-0.1386	0.0311	-1.2427	0.214	36949
condition:crossbmiz:delta_bmiz	-0.0236	0.0441	-0.11	0.0629	-0.5342	0.593	36949
pfb:crossbmiz:delta_bmiz	-0.0425	0.0751	-0.1898	0.1048	-0.5656	0.572	36949
prepost_rev:crossbmiz:delta_bmiz	0.0106	0.0356	-0.0592	0.0804	0.2968	0.767	36949
group_OB:condition:pfb:prepost_rev	-0.0668	0.0449	-0.1548	0.0213	-1.4856	0.137	36949
condition:pfb:prepost_rev:crossbmiz	-0.0221	0.0232	-0.0676	0.0233	-0.9545	0.34	36949
condition:pfb:prepost_rev:delta_bmiz	-0.0325	0.0353	-0.1016	0.0366	-0.9213	0.357	36949
condition:pfb:crossbmiz:delta_bmiz	-0.0049	0.0343	-0.0721	0.0622	-0.1442	0.885	36949
condition:prepost_rev:crossbmiz:delta_bmiz	-0.0311	0.0285	-0.087	0.0248	-1.0893	0.276	36949
pfb:prepost_rev:crossbmiz:delta_bmiz	0.0029	0.0359	-0.0675	0.0733	0.0797	0.937	36949
condition:pfb:prepost_rev:crossbmiz:delta_bmiz	-0.0111	0.0285	-0.0669	0.0446	-0.3915	0.695	36949

4.2.3. Perseveration

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	-2.2737	0.1025	-2.4746	-2.0728	-22.1841	0	12309
group_BED	0.118	0.1431	-0.1626	0.3985	0.8241	0.41	12309
group_Lean	0.0782	0.143	-0.2022	0.3586	0.5467	0.585	12309
condition	0.2143	0.0691	0.0788	0.3497	3.1014	0.002	12309
prepost_rev	0.275	0.0638	0.15	0.3999	4.3119	0	12309

group_BED:condition	-0.1888	0.0955	-0.3759	-0.0017	-1.9775	0.048	12309
group_Lean:condition	-0.0699	0.0958	-0.2577	0.118	-0.7289	0.466	12309
group_BED:prepost_rev	-0.1234	0.0878	-0.2954	0.0487	-1.4051	0.16	12309
group_Lean:prepost_rev	-0.1245	0.0877	-0.2963	0.0474	-1.4195	0.156	12309
condition:prepost_rev	-0.1066	0.0596	-0.2234	0.0102	-1.789	0.074	12309
group_BED:condition:prepost_rev	0.098	0.0817	-0.0622	0.2582	1.1991	0.231	12309
group_Lean:condition:prepost_rev	0.0987	0.0817	-0.0616	0.2589	1.207	0.228	12309

4.2.3.1. Effect of binge eating frequency

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	-2.1038	0.0866	-2.2735	-1.934	-24.2955	0	3849
condition	-0.0115	0.0695	-0.1478	0.125	-0.1652	0.869	3849
prepost_rev	0.1919	0.0678	0.059	0.325	2.8314	0.005	3849
crossbingesz	-0.1571	0.1484	-0.4481	0.134	-1.059	0.29	3849
delta_bez	0.0521	0.2083	-0.3563	0.461	0.2503	0.802	3849
condition:prepost_rev	0.0235	0.0642	-0.1023	0.149	0.3663	0.714	3849
condition:crossbingesz	-0.0714	0.1229	-0.3124	0.17	-0.5813	0.561	3849
prepost_rev:crossbingesz	0.2279	0.1154	0.0017	0.454	1.9754	0.048	3849
condition:delta_bez	-0.0602	0.1599	-0.3738	0.253	-0.3762	0.707	3849
prepost_rev:delta_bez	-0.1525	0.1722	-0.4901	0.185	-0.8855	0.376	3849
crossbingesz:delta_bez	-0.1465	0.4927	-1.1124	0.82	-0.2972	0.766	3849
condition:prepost_rev:crossbingesz	-0.0075	0.1132	-0.2294	0.215	-0.0661	0.947	3849
condition:prepost_rev:delta_bez	0.1739	0.1472	-0.1147	0.463	1.1815	0.238	3849
condition:crossbingesz:delta_bez	0.4999	0.4103	-0.3045	1.304	1.2183	0.223	3849
prepost_rev:crossbingesz:delta_bez	0.0819	0.4287	-0.7586	0.922	0.191	0.849	3849
condition:prepost_rev:crossbingesz:delta_bez	-0.5898	0.3808	-1.3364	0.157	-1.549	0.122	3849

4.2.3.2. Effect of BMI

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	-2.0845	0.0906	-2.2622	-1.9069	-22.9996	0	7685
group_OB	-0.1784	0.1278	-0.4288	0.0721	-1.396	0.163	7685
condition	0.0792	0.0604	-0.0391	0.1975	1.3122	0.19	7685
prepost_rev	0.0921	0.0601	-0.0257	0.2099	1.5331	0.125	7685
crossbmiz	0.0064	0.0686	-0.1281	0.1409	0.0934	0.926	7685
delta_bmiz	-0.0082	0.1064	-0.2169	0.2004	-0.0774	0.938	7685
group_OB:condition	0.1651	0.0877	-0.0069	0.337	1.8812	0.06	7685
group_OB:prepost_rev	0.1549	0.0873	-0.0162	0.326	1.7748	0.076	7685
condition:prepost_rev	-0.0293	0.0578	-0.1426	0.0841	-0.5064	0.613	7685
condition:crossbmiz	0.0885	0.0468	-0.0032	0.1801	1.8922	0.059	7685
prepost_rev:crossbmiz	-0.0564	0.0456	-0.1458	0.0331	-1.2355	0.217	7685
condition:delta_bmiz	0.0331	0.0758	-0.1155	0.1817	0.4365	0.663	7685

prepost_rev:delta_bmiz	-0.0863	0.0746	-0.2325	0.06	-1.1564	0.248	7685
crossbmiz:delta_bmiz	-0.0627	0.0935	-0.2461	0.1207	-0.6701	0.503	7685
group_OB:condition:prepost_rev	-0.1166	0.0838	-0.2808	0.0477	-1.391	0.164	7685
condition:prepost_rev:crossbmiz	-0.0491	0.0447	-0.1367	0.0384	-1.1003	0.271	7685

4.3. Modelling results including subjects without MRI

4.3.1. Reinforcement sensitivities

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	1.5343	0.1062	1.3257	1.7429	14.454	0	504
Feedback	0.8564	0.0514	0.7553	0.9574	16.646	0	504
Condition	-0.0897	0.0191	-0.1273	-0.0521	-4.686	0	504
Group_Lean	0.0884	0.1501	-0.2066	0.3833	0.5886	0.556	504
Group_BED	-0.0403	0.1501	-0.3353	0.2546	-0.2686	0.788	504
Feedback:Condition	0.052	0.0141	0.0242	0.0797	3.6821	3E-04	504
Feedback:Group_Lean	0.0568	0.0728	-0.0862	0.1997	0.7804	0.436	504
Feedback:Group_BED	0.0209	0.0728	-0.122	0.1638	0.2873	0.774	504
Condition:Group_Lean	0.0251	0.0271	-0.0281	0.0782	0.9258	0.355	504
Condition:Group_BED	0.0451	0.0271	-0.008	0.0983	1.6678	0.096	504
Feedback:Condition:Group_Lean	0.0131	0.02	-0.0261	0.0523	0.6568	0.512	504
Feedback:Condition:Group_BED	0.0081	0.02	-0.0311	0.0473	0.4058	0.685	504

4.3.2. Learning rates

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	0.6557	0.0171	0.6221	0.689	38.413	0	504
Feedback	0.1541	0.0059	0.1425	0.166	26.177	0	504
Condition	0.0115	0.0032	0.0051	0.018	3.5545	4E-04	504
Group_Lean	-0.0053	0.0241	-0.0527	0.042	-0.2195	0.826	504
Group_BED	-0.0019	0.0241	-0.0493	0.046	-0.0768	0.939	504
Feedback:Condition	0.0003	0.0029	-0.0054	0.006	0.1084	0.914	504
Feedback:Group_Lean	0.0118	0.0083	-0.0046	0.028	1.4153	0.158	504
Feedback:Group_BED	0.0169	0.0083	0.0005	0.033	2.0291	0.043	504
Condition:Group_Lean	0.0064	0.0046	-0.0026	0.015	1.3919	0.165	504
Condition:Group_BED	0.0059	0.0046	-0.0031	0.015	1.283	0.2	504

Feedback:Condition:Group_Lean	0.0039	0.0041	-0.0042	0.012	0.9404	0.348	504
Feedback:Condition:Group_BED	0.0047	0.0041	-0.0035	0.013	1.1259	0.261	504

4.3.3. *Double update learning rates (α*κ)*

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	0.2962	0.0143	0.2681	0.324	20.721	0	504
Feedback	0.0661	0.0032	0.0598	0.073	20.357	0	504
Condition	0.0086	0.006	-0.0032	0.02	1.4306	0.153	504
Group_Lean	-0.0023	0.0202	-0.042	0.037	-0.1139	0.909	504
Group_BED	-0.028	0.0202	-0.0678	0.012	-1.3875	0.166	504
Feedback:Condition	0.0006	0.0013	-0.0019	0.003	0.4888	0.625	504
Feedback:Group_Lean	0.0066	0.0046	-0.0024	0.016	1.4442	0.149	504
Feedback:Group_BED	0.0045	0.0046	-0.0045	0.014	0.9889	0.323	504
Condition:Group_Lean	0.0111	0.0085	-0.0056	0.028	1.3077	0.192	504
Condition:Group_BED	-0.0032	0.0085	-0.0199	0.014	-0.3714	0.711	504
Feedback:Condition:Group_Lean	0.0023	0.0018	-0.0012	0.006	1.2923	0.197	504
Feedback:Condition:Group_BED	0.0008	0.0018	-0.0028	0.004	0.4255	0.671	504

4.3.4. *Learning sensitivity (α *ρ)*

Predictors	*Estimates*	*Std Error*	*Lower*	*Upper*	*t*	*p*	*df*
(Intercept)	1.1221	0.0728	0.9791	1.265	15.418	0	504
Feedback	0.8011	0.0485	0.7058	0.896	16.522	0	504
Condition	-0.0277	0.0154	-0.0579	0.003	-1.802	0.072	504
Group_Lean	0.0618	0.1029	-0.1404	0.264	0.6005	0.548	504
Group_BED	-0.0313	0.1029	-0.2335	0.171	-0.3038	0.761	504
Feedback:Condition	0.0466	0.014	0.0192	0.074	3.3381	9E-04	504
Feedback:Group_Lean	0.0802	0.0686	-0.0545	0.215	1.1699	0.243	504
Feedback:Group_BED	0.0347	0.0686	-0.1	0.169	0.5063	0.613	504
Condition:Group_Lean	0.0352	0.0218	-0.0075	0.078	1.6191	0.106	504
Condition:Group_BED	0.0503	0.0218	0.0075	0.093	2.3101	0.021	504
Feedback:Condition:Group_Lean	0.0139	0.0197	-0.0249	0.053	0.7029	0.483	504
Feedback:Condition:Group_BED	0.0172	0.0197	-0.0216	0.056	0.8713	0.384	504

4.3.5. *Figures*

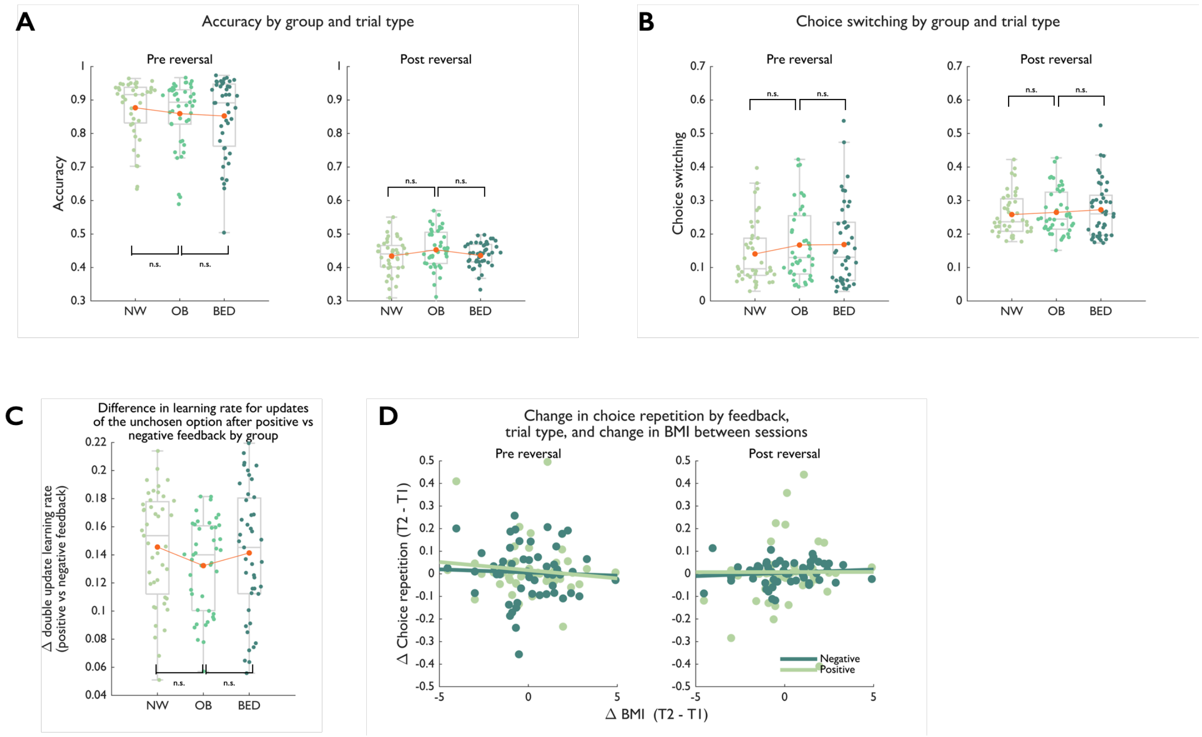

Fig. S10. – Sample including subjects without MRI. A. Accuracy in pre and post reversal trials by group. B. Choice switching in pre and post reversal trials by group. C. Difference in double update (counterfactual) learning rate for positive and negative feedback by group. D. Change in choice switching by feedback (positive vs negative), trial type (pre vs post reversal), and change in BMI between sessions.

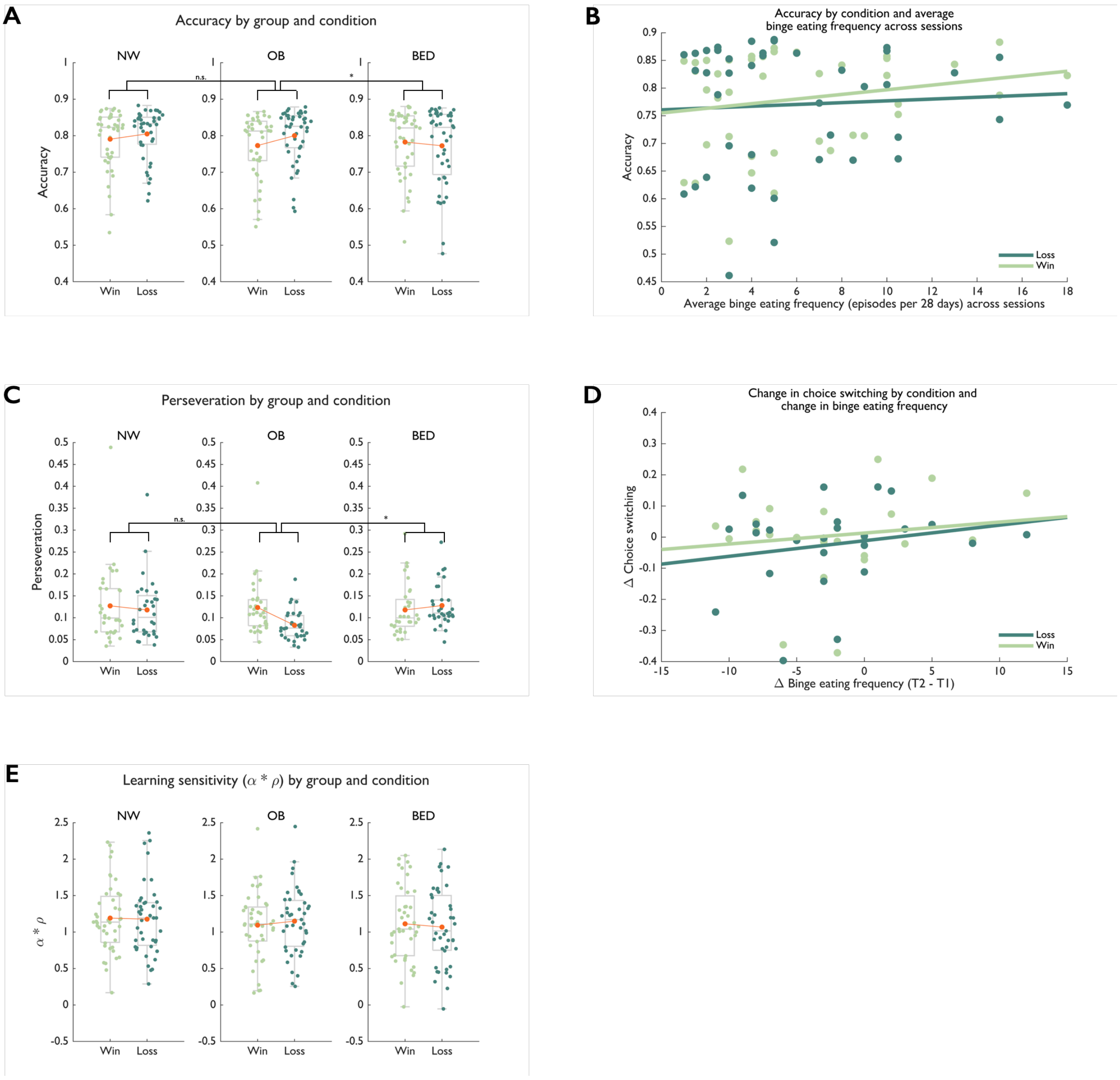

Fig. S11. – Sample including subjects without MRI. A. Accuracy by condition and group. B. Accuracy by condition and average binge eating frequency (episodes per 28 days) within the BED group. C. Perseveration by condition and group. D. Change in choice switching by condition and change in binge eating frequency across sessions. E. “Learning sensitivity”, the product of learning rate and reinforcement sensitivity, by condition and group.

5. Supplementary References

Daw, N. D., O’Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

den Ouden, H. E. M., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., Franke, B., & Cools, R. (2013). Dissociable Effects of Dopamine and Serotonin on Reversal Learning. *Neuron*, *80*(4), 1090–1100. https://doi.org/10.1016/J.NEURON.2013.08.030

Hilbert, A., Tuschen-Caffier, B., & Ohms, M. (2004). Eating Disorder Examination: Deutschsprachige Version des strukturierten Essstörungsinterviews. *Diagnostica*, *50*(2), 98–106. https://doi.org/10.1026/0012-1924.50.2.98

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLOS Computational Biology*, *7*(4), e1002028. https://doi.org/10.1371/journal.pcbi.1002028

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLOS Computational Biology*, *8*(3), e1002410. https://doi.org/10.1371/journal.pcbi.1002410

Huys, Q. J. M., & Schad, D. (2015). *No Title*. Emfit Matlab Script. https://github.com/mpc-ucl/emfit

Katahira, K., & Toyama, A. (2021). Revisiting the importance of model fitting for model-based fMRI: It does matter in computational psychiatry. *PLOS Computational Biology*, *17*(2), e1008738. https://doi.org/10.1371/journal.pcbi.1008738

Reiter, A. M. F., Heinze, H. J., Schlagenhauf, F., & Deserno, L. (2017). Impaired Flexible Reward-Based Decision-Making in Binge Eating Disorder: Evidence from Computational Modeling and Functional Neuroimaging. *Neuropsychopharmacology*, *42*(3), 628–637. https://doi.org/10.1038/npp.2016.95

Schlagenhauf, F., Huys, Q. J. M., Deserno, L., Rapp, M. A., Beck, A., Heinze, H.-J., Dolan, R., & Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *NeuroImage*, *89*, 171–180. https://doi.org/https://doi.org/10.1016/j.neuroimage.2013.11.034

Waltmann, M., Herzog, N., Reiter, A. M. F., Villringer, A., Horstmann, A., & Deserno, L. (2023). Diminished reinforcement sensitivity in adolescence is associated with enhanced response switching and reduced coding of choice probability in the medial frontal pole. *Developmental Cognitive Neuroscience*, *60*, 101226.

https://doi.org/10.1016/j.dcn.2023.101226

Waltmann, M., Schlagenhauf, F., & Deserno, L. (2022). Sufficient reliability of the behavioral and computational readouts of a probabilistic reversal learning task. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01739-7

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3), 279–292. https://doi.org/10.1007/BF00992698

Wittchen, H.-U. (1997). *Strukturiertes klinisches Interview für DSM-IV: SKID. Achse I: Psychische Störungen: Interviewheft und Beurteilungsheft; eine deutschsprachige, erweiterte Bearbeitung der amerikanischen Originalversion des SCID-I*. Hogrefe, Verlag für Psychologie.