ARTICLE

Check for updates

# State-specific alterations in the neural computations underlying inhibitory control in women remitted from bulimia nervosa

Laura A. Berner[1]✉, Katia M. Harlé[2,3], Alan N. Simmons[2,3], Angela Yu[2,4], Martin P. Paulus[2,5], Amanda Bischoff-Grethe[2], Christina E. Wierenga[2,3], Ursula F. Bailer[2,6] and Walter H. Kaye[2]



The neurocomputational processes underlying bulimia nervosa and its primary symptoms, out-of-control overeating and purging, are poorly understood. Research suggests that the brains of healthy individuals form a dynamic internal model to predict whether control is needed in each moment. This study tested the hypothesis that this computational process of inhibitory control is abnormally affected by metabolic state (being fasted or fed) in bulimia nervosa. A Bayesian ideal observer model was fit to behavioral data acquired from 22 women remitted from bulimia nervosa and 20 group-matched controls who completed a stop-signal task during two counterbalanced functional MRI sessions, one after a 16 h fast and one after a meal. This model estimates participants' trial-by-trial updating of the probability of a stop signal based on their experienced trial history. Neural analyses focused on control-related Bayesian prediction errors, which quantify the direction and degree of "surprise" an individual experiences on any given trial. Regardless of group, metabolic state did not affect behavioral performance on the task. However, metabolic state modulated group differences in neural activation. In the fed state, women remitted from bulimia nervosa had attenuated prediction-error-dependent activation in the left dorsal caudate. This fed-state activation was lower among women with more frequent past binge eating and self-induced vomiting. When they are in a fed state, individuals with bulimia nervosa may not effectively process unexpected information needed to engage inhibitory control. This may explain the difficulties these individuals have stopping eating after it begins.

*Molecular Psychiatry* (2023) 28:3055–3062; https://doi.org/10.1038/s41380-023-02063-6

## INTRODUCTION

Bulimia nervosa is the second most prevalent eating disorder [1] and is one of the most common psychiatric conditions in women [2]. Mortality rates are significantly elevated in bulimia nervosa, and symptoms persist in over 60% of patients who receive first-line treatments [3, 4]. However, the neurobiological mechanisms that drive the core symptoms of out-of-control overconsumption of food (i.e., binge eating) and subsequent compensatory behaviors, like self-induced vomiting, remain poorly understood. Data from healthy adults suggest that the regulation of food intake is influenced by inhibitory-control-related processes and by metabolic state—whether one has been recently fed or fasted [5]. Individuals with bulimia nervosa alternate between extremes of overcontrolled intake, including dietary restriction and fasting, and disinhibited intake, including binge eating, suggesting a potentially aberrant interaction between cognitive control and metabolic factors that has not yet been specifically examined.

Both adults and adolescents with bulimia nervosa show moderately elevated inhibitory control error rates [6] and reduced activation in frontostriatal regions during response inhibition [7–9] compared with healthy individuals. However, our understanding of the cognitive neuroscience of inhibitory control is rapidly evolving. Computational models frame inhibitory control as an adaptive process that requires moment-to-moment decision-making about whether to engage in or stop a behavior, and using experience to update beliefs about how likely it is that inhibition will be required in the subsequent moment [10]. Integrating computational models with behavioral and neural data through model-based functional magnetic resonance imaging (fMRI) enables more precise and quantitative characterization of this inhibitory-control-related predictive processing, decision-making, and updating [11]. To date, model-based fMRI has been used to demonstrate a likely role for altered reward-based learning and prediction errors in bulimia nervosa; [12] however, the potential contribution of alterations in the dynamic predictive processing underlying inhibition to bulimic symptoms remains unknown. In addition, alterations in inhibitory control may be influenced by metabolic state, as some data from healthy adults suggest reductions in control after acute fasting [13–17]. In bulimia nervosa, it is still unclear whether disturbances in inhibitory control may be exaggerated in a particular metabolic state.

To identify more precise components of the inhibition process that might underlie bulimic symptoms, we conducted a model-

[1]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Department of Psychiatry, University of California, San Diego, CA, USA. [3]VA San Diego Healthcare System, San Diego, CA, USA. [4]Centre for Cognitive Science & Hessian AI Center, Technical University of Darmstadt, Darmstadt, Germany. [5]Laureate Institute for Brain Research, Tulsa, OK, USA. [6]Department of Psychiatry and Psychotherapy, Division of Biological Psychiatry, Medical University of Vienna, Vienna, Austria.
✉email: laura.berner@mssm.edu

Received: 10 November 2022 Revised: 20 March 2023 Accepted: 28 March 2023
Published online: 27 April 2023

based analysis of fMRI data collected from healthy women and women with past bulimia nervosa who were scanned twice: after an acute period of fasting and after eating. We studied participants in remission from bulimia nervosa to avoid the potentially confounding effects of electrolyte disturbances or recent extreme and symptom-related fluctuations in metabolic state in currently symptomatic individuals. Despite symptom remission, our prior work has detected altered neural responses among women with a history of bulimia nervosa [18–21].

We used a well-characterized Bayesian ideal observer model to understand adaptive inhibitory control on a stop signal task (SST). This model was originally developed to capture behavior in healthy individuals [10, 22], but it has since been used to identify neural correlates and predictors of clinical outcome in substance use disorders [23–27]. Substance use disorders and bulimia nervosa are characterized by intermittent bouts of maladaptive behavior, deficits in cognitive and behavioral control, and frontostriatal dysfunction [28, 29]. Given these similarities, the same computational neuroimaging approach may be useful for both conditions. We focused our analyses on control-related Bayesian prediction errors that have been repeatedly implicated in the development and persistence of substance use disorders [23–26]. The first type, unsigned prediction errors (UPE), capture the absolute magnitude of discrepancy, without directionality, between the predicted probability that inhibition would be required (P(stop)) and the actual occurrence or non-occurrence of this demand. A large UPE can be understood as an individual's "surprise" that control was or was not needed. It indicates that the individual's model of the environment is less accurate, and that the need to stop is more difficult to predict. The overall expectancy violation tracked by UPE may play a central role in preparing to adaptively switch to a different behavioral strategy when changes in internal or external states occur [30]. Because fasting and eating involve dramatic state changes, appropriate tracking via UPEs may be critical for normalized eating. The second type of prediction errors, signed prediction errors (SPE), indicate whether the need to stop was unexpectedly present [i.e., 1 - the predicted probability that inhibition was required (P(stop)) = positive SPE] or unexpectedly absent [i.e., 0 - P(stop) = negative SPE], thereby influencing directional belief updating and model updating. SPE activation may help individuals predict specific appropriate actions. Both UPEs and SPEs play important roles in learning (e.g., [31]).

We anticipated that like individuals with addiction, women remitted from bulimia nervosa would show altered activation associated with UPEs, specifically in the dorsal caudate and lateral prefrontal cortices [26]. These regions have been linked consistently with reward and control-related prediction errors in food and non-food-specific paradigms [32–35]. Such findings would suggest that bulimia nervosa is associated with aberrant processing of the discrepancy between expected and observed control-related demands. However, we hypothesized that in remitted bulimia nervosa, these alterations would be modulated by the fasted or fed internal state of the individual. Because SPE activation has also been linked to substance-use risk and relapse [23, 25], additional analyses examined potential state-dependent alterations in SPE activation. Finally, we explored group × state effects on neural activation associated with control-related expectations, regardless of trial type or accuracy (i.e., P(stop) activation).

## MATERIALS AND METHODS

### Participants

Data from 22 women remitted from *DSM-IV-TR* bulimia nervosa, purging subtype [36] and 20 controls were included in analyses (see data exclusion details in Supplement). Remission was defined by maintaining above 85% of ideal body weight, regular menstrual cycles, and abstinence from binge eating, purging, and restrictive eating for at least 1 year [37]. Women who met criteria for a current Axis I diagnosis were excluded from both groups, and controls with any eating disorder history were excluded. The study was approved by the University of California, San Diego (UCSD) Institutional Review Board, and all participants provided written informed consent (see the Supplement for additional details regarding participants and assessment tools).

### Procedure

Participants performed a SST [23–26, 38–42] during fMRI on two counterbalanced visits scheduled 24 h apart and in the early follicular menstrual phase. Participants stayed at the UCSD Clinical & Translational Research Institute for 72 h during the study, and all meals were provided. Imaging data were acquired on one of two 3 T scanners at 9:00 AM on both scan days, and each participant was scanned on the same scanner for both scans (see Supplement). On the fasted-state scan day, participants fasted for 16 h before scanning. On the fed-state scan day, participants consumed a standardized breakfast (30% of overall daily caloric needs, calculated as 30 kcal/kg body weight, ~450–500 kcal; 53% carbohydrates, 32% fat, and 15% protein) at 7 AM (see Supplement for additional detail).

During the SST, participants were instructed to press a button in response to Xs and Os (go stimuli) but to withhold responding when they heard a tone (stop signal). To discourage participants from waiting for stop signals to occur, scripted instructions told participants, "If you wait too long and press the button after the maximum time on go trials, your response will be counted as an error, and the next trial will automatically start immediately after." Participants' mean reaction time (RT) on go trials from a brief version of the task completed before scanning on both days was used to individually calibrate the stop-signal delay across six levels of difficulty (Fig. 1A; 23, 24–26, 39–43). This design permits a variable stop-signal delay that is not predictable to participants (see Supplement for additional SST design rationale). Likert-type scales (0–7) assessed hunger and thirst before and after scanning (Fig. 2).

### Statistical analysis

*Model-agnostic behavioral analyses*. Hierarchical generalized linear mixed-effects models (LMEs) with subject as a random effect and a logit-link function examined effects of group, metabolic state, stop-signal delay, and their interaction on the trial-wise likelihood of a stop-trial error vs. correct inhibition. Hierarchical LMEs tested effects of P(stop), group, state, and their interactions on go-trial RT and explored effects on stop signal reaction time (SSRT; see Supplement for calculation) and post-error slowing.

*Computational modeling*. SST data were analyzed with a previously used Bayesian ideal observer model to generate computational regressors for fMRI analysis [22–26]. This approach provides a way to infer latent cognitive variables associated with individuals' inhibitory control performance, including the trial-by-trial predicted probability for the need to inhibit (P(stop)) and the associated Bayesian control prediction errors (Fig. 1B; see Supplement for details). Given our model parameters and the pseudo-randomized trial sequence experienced by all participants, we computed the corresponding sequence of P(stop) values, as well as UPE values [i.e., |outcome- P(stop)|] and SPE values [i.e., outcome – P(stop)] on each trial.

### MRI analyses

Functional images were preprocessed and analyzed using Analysis of Functional NeuroImages (AFNI) software (http://afni.nimh.nih.gov/afni/) and FSL (http://fsl.fmrib.ox.ac.uk/fsl/).Group-level analyses were performed using the *lme4*, *oro.nifti*, *MASS*, and *abind* packages in R (http://www.r-project.org; see Supplement). Our analytic approach mirrored that of papers using the SST and the described computational model for model-based fMRI analyses in substance use disorders [23–25].

### First-level fMRI analyses

All regressors were convolved with a canonical hemodynamic response function (see Supplement for regressors of no interest included in each model below).

*Bayesian expectancy violations (prediction errors)*. Primary first-level general linear models (GLMs) assessed neural activation modulated by Bayesian prediction errors. Given high collinearity between Bayesian UPE

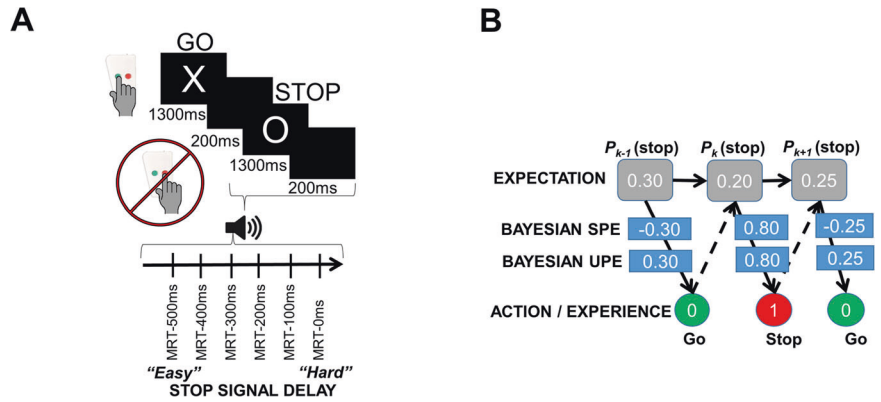


**Fig. 1 Schematic representations of the stop signal task and Bayesian ideal observer model used in the current study. A** Participants completed 288 total trials, including 72 stop trials, divided into six blocks of 48 trials (with 25% stop trials per block). Each trial was 1300 ms separated by a 200 ms blank-screen interstimulus interval. Trial order was pseudorandomized. Before scanning on each day, participants performed a brief version of the task to individually calibrate the stop-signal delay across six levels of difficulty. Stop signals occurred between 0 and 500 ms before the participant's mean reaction time (MRT) in intervals of 100 ms (stop signals presented closer to a participant's MRT are more challenging). Go stimuli disappeared from the screen when participants responded. This provided a subject-specific jittered reference function. Participants were instructed to press the left button in response to X and the right button in response to O, but to avoid pressing either button if they heard a tone during a trial (the stop signal). **B** The computational model computes trial-by-trial estimates about the predicted likelihood of needing to stop. The model assumes that experiencing stop trials increases the expected likelihood of encountering a stop trial (P(stop)), and experiencing go trials decreases P(stop). Changes in P(stop) update the individual's decision-making policy such that the "optimal responder" slows their reaction time to go trials in a linear fashion as P(stop) increases to increase the likelihood of correctly stopping on a stop trial [22]. The expected probability of needing to inhibit on trial $k$ ($P_k$(stop)) is compared with the experienced outcome (0 = go, 1 = stop). This comparison generates a prediction error (signed prediction error = outcome – P(stop), unsigned prediction error = |outcome – P(stop)| as displayed in blue boxes). This prediction error is combined with the prior to produce an updated prior for the subsequent trial.

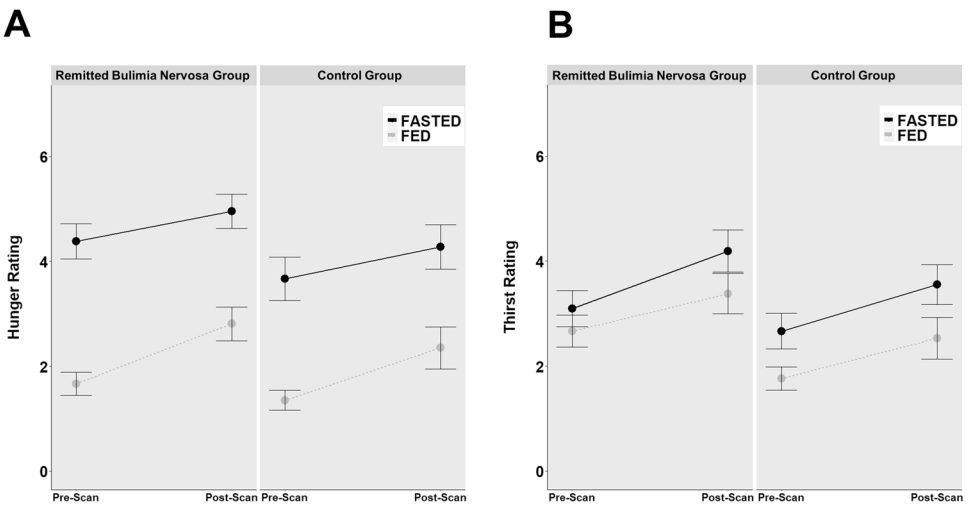


**Fig. 2 Hunger and thirst ratings before and after scanning. A** A main effect of state in a group × state (fasted, fed) × time point (pre-scan, post-scan) linear mixed-effects model, with subject as a random effect, indicated that all participants reported higher levels of hunger during the fasted relative to the fed state ($B = 3.86$, SE = 1.16, $p = 0.001$). There were no other main effects or interactions. **B** For thirst, there was only a main effect of time point, such that in both states, all participants reported higher levels of thirst at the post-scan relative to the pre-scan assessment ($B = 1.10$, SE = 0.29, $p = 0.0002$). There were no other main effects or interactions that were statistically significant.

(|outcome-P(stop)|) and SPE (outcome-P(stop); VIFs > 20), we ran separate models for each type of prediction error. GLM 1 (the UPE model) included trial-by-trial UPE as the regressor of interest. SPE residualized with respect to UPE was included a regressor of no interest in GLM 1. GLM 2 (the SPE model) included SPE as the regressor of interest and UPE residualized with respect to SPE as a regressor of no interest (see Supplementary Methods for further detail).

*Prediction of inhibitory demand (P(stop)) and model-agnostic inhibitory control.* As in prior model-based fMRI analyses in substance use disorders [23–25], to explore activation associated with P(stop) across trial types and controlling for response accuracy, GLM 3 included categorical trial types (i.e., go, successful stop, failed stop) as well as each trial type parametrically modulated by P(stop) (i.e., go × P(stop), successful stop × P(stop), and failed stop × P(stop)).

## Group-level fMRI analyses

*Bayesian expectancy violations (prediction errors).* Neural group × state interactions were tested at the voxel-wise level in R. Our primary group-level analyses were voxel-wise, whole-brain LMEs to test for group × state (fasted or fed) interaction effects on neural activation associated with prediction errors (from GLMs 1 and 2). A UPE is equal to the absolute value of 0- P(stop) (i.e., P(stop)) on go trials, and 1-P(stop) on stop trials. Therefore, to truly encode UPEs, a cluster's activation should be positively correlated with mean P(stop) activation on go trials and negatively

**Table 1.** Participant characteristics.

| Characteristic | Healthy control group ($n = 20$) | | Remitted bulimia nervosa group ($n = 22$) | | |
|---|---|---|---|---|---|
| | *N* | % | *N* | % | *p* |
| Scanner | | | | | 0.385 |
| GE Signa Excite | 10 | 50.0 | 11 | 45.5 | |
| GE MR 750 | 10 | 50.0 | 11 | 45.5 | |
| Self-reported Race | | | | | 0.493 |
| Hispanic | 2 | 10.0 | 1 | 4.5 | |
| Self-reported Ethnicity | | | | | 0.521 |
| White | 13 | | 16 | | |
| Asian | 6 | | 3 | | |
| Black/African American | 1 | | 1 | | |
| Pacific Islander | 0 | | 1 | | |
| American Indian/ Alaska Native | 0 | | 0 | | |
| Other | 0 | | 1 | | |
| | Mean | SD | Mean | SD | *p* |
| Age (years) | 25.3 | 5.0 | 28.3 | 5.6 | 0.039 |
| BMI (kg/m$^2$) | 22.2 | 2.2 | 22.6 | 1.8 | 0.743 |
| Duration of illness (months) | | | 73.0 | 47.1 | |
| Duration of remission (months) | | | 50.5 | 45.5 | |
| Education (years) | 15.9 | 1.3 | 16.6 | 2.5 | 0.059 |
| FSIQ | 111.1 | 10.3 | 114.8 | 7.8 | 0.266 |
| | N | % | N | % | |
| Past Diagnoses | | | | | |
| Major Depressive Disorder | 0 | 0 | 12 | 54.5 | |
| Any Anxiety Disorder | 0 | 0 | 2 | 9.1 | |
| Substance Use Disorder | 0 | 0 | 3 | 13.6 | |
| Alcohol Use Disorder | 0 | 0 | 5 | 22.7 | |
| Anorexia Nervosa[a] | 0 | 0 | 12 | 54.5 | |
| | Mean | SD | Mean | SD | *p* |
| Self-Reported Symptoms[b] | | | | | |
| Trait Anxiety (STAI) | 23.6 | 3.4 | 27.4 | 7.8 | 0.548 |
| Depression (BDI) | 0.3 | 0.5 | 1.9 | 2.8 | 0.012 |

*BMI* body mass index, *FSIQ* full-scale IQ estimated from the Wechsler Abbreviated Scale of Intelligence Full-Scale IQ, *STAI* Spielberger State-Trait Anxiety Inventory, *BDI* Beck Depression Inventory.
[a]Bulimia nervosa was the most recent eating disorder diagnosis for all women remitted from bulimia nervosa.
[b]One healthy control did not complete clinical assessments.

correlated with mean P(stop) activation on stop trials. Areas showing deactivations associated with UPE (i.e., negative UPE) should show the opposite P(stop) activation pattern—negative correlations with P(stop) on go trials and positive correlations with P(stop) on stop trials. In contrast, SPE is equal to 0 - P(stop) = − P(stop) on go trials, and 1- P(stop) on stop trials. Therefore, clusters showing positive SPE activation should show activation negatively correlated with P(stop) on both go and stop trials. Areas showing deactivations associated with SPE (i.e., negative SPE) should show the opposite pattern—positive correlations with P(stop) on both go and stop trials. Therefore, to confirm their consistency with expected patterns, we extracted and plotted mean P(stop)-modulated go- and stop-trial activation (from GLM 3) from clusters that showed group x state interactions for UPE or SPE [22–26]. Consistent with all the prior imaging studies that have used this SST in individuals with or at risk for substance use disorders and individuals with posttraumatic stress disorder [22–26, 41], we interpret results in clusters where the P(stop) activation pattern was consistent with activation from the prediction-error model across groups and states (see Supplement for further detail).

*Bayesian prediction of inhibitory demand (P(stop))*. To explore state-dependent group differences in P(stop) activation, as in prior work [23–26], we conducted one LME that assessed for the significance of group × state interaction effects on P(stop) activation while controlling for trial type (i.e., group × state + trial type (go or stop)), and one LME that assessed for the significance of group × state interaction effects on P(stop) activation while controlling for accuracy (i.e., group × state + stop accuracy (successful or failed stop)).

All group-level LMEs included scanner as a within-subjects random effect. Intrinsic smoothness was estimated using the spatial autocorrelation function (ACF) option in AFNI's 3dFWHMx. Minimum cluster sizes were calculated with 3dClustSim to guard against false positives across the whole brain (voxel-wise $p < 0.001$, corrected for multiple comparisons at familywise error rate $p < 0.05$, bi-sided with first nearest neighbor clustering requiring voxel faces to touch; minimum cluster size = 6 voxels, 162 mm$^3$) [43]. The normality and homoscedasticity of residuals were visually checked.

### Association with clinical variables

Within the remitted bulimia nervosa group, exploratory negative binomial regressions examined associations of mean activation tracking prediction errors in each state and worst past binge eating and self-induced vomiting frequency. All participants in the remitted bulimia nervosa group had a history of purging behavior, but our method of past symptom assessment precluded the calculation of a composite score representing worst past total purging frequency per week. Since all but three of the participants engaged in self-induced vomiting, we examined self-induced vomiting as a clinical correlate to replicate the approaches of the limited prior neuroimaging research focused on inhibitory control in bulimia nervosa [8, 9]. Huber robust regressions tested associations of mean activation tracking prediction errors in each state with durations of bulimia nervosa and remission. Alpha was set at 0.006 to Bonferroni adjust for the number of tests.

Post-hoc exploratory analyses examined the potential impact of past comorbidities, lowest BMI, and current depressive and anxiety symptoms on our findings (see Supplement).

## RESULTS

Participants in the two groups did not differ on BMI, IQ, or education (Table 1). Since women remitted from bulimia nervosa were older than controls, between-group analyses included the fixed main effect of age. Both groups reported greater hunger during the fasted state compared with the fed state but groups did not differ on ratings of hunger or thirst (Fig. 2).

### Task performance

*Accuracy*. All participants had a higher likelihood of error on stop trials with a longer stop-signal delay ($\chi^2_1 = 1754.7$, $p < 0.001$), and higher P(stop) estimates were associated with lower stop error likelihood ($\chi^2_1 = 3.95$, $p = 0.047$). Adding a fixed effect of state slightly improved model fit ($\chi^2_1 = 6.10$, $p = 0.014$), but there was no state × stop-signal delay interaction, no additional fixed effect of group or age, and no interactions of group with age, stop-signal delay, P(stop), or state ($ps > 0.414$).

*Reaction time*. There was a positive association between go RTs and trial-wise P(stop) model estimates across all participants

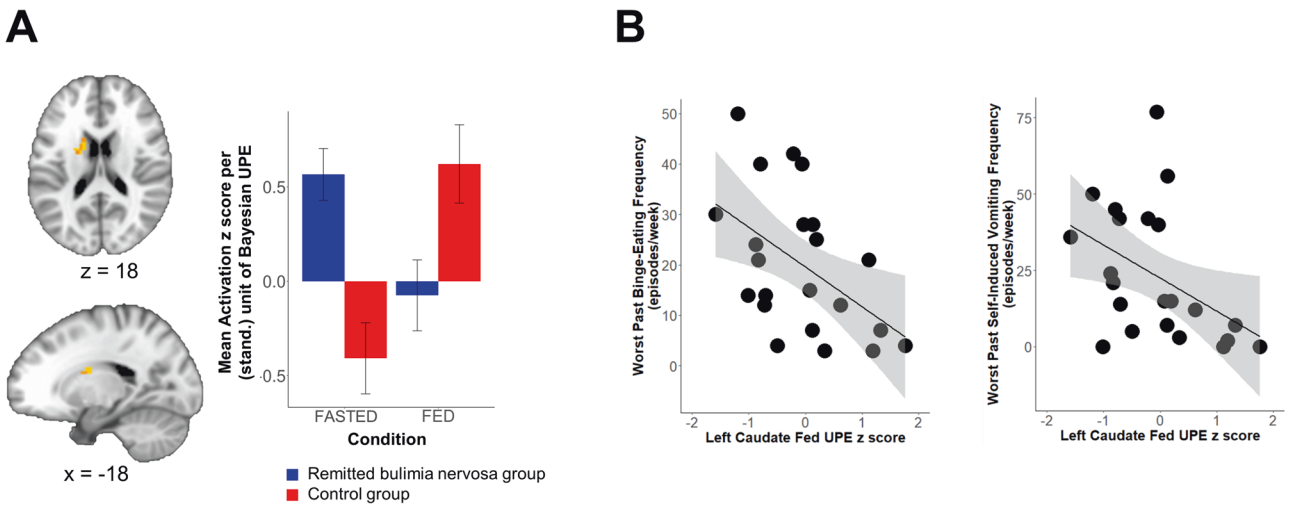


**Fig. 3 Group differences in the influence of eating on neural tracking of inhibitory control-related expectancy violation. A** Neural activation in the left dorsal caudate associated with Bayesian unsigned prediction errors (UPE) depended on an interaction between group and metabolic state (voxel-wise $p < 0.001$, corrected for multiple comparisons at familywise error rate $p = 0.05$; two-tailed cluster-based correction at the whole-brain level). Plot bars represent mean activation associated with UPE in the fasted and fed state. Error bars indicate standard error of the mean. **B** Negative binomial regressions within the remitted bulimia nervosa group indicated that lower fed-state activation tracking UPE in the left dorsal caudate was associated with more frequent past binge eating and self-induced vomiting.

($\chi^2_1 = 66.60$, $p < 0.001$), and within each group across states ($ps < 0.001$; Fig. S1). The model of the association between P(stop) and RT that best fit the data included a main effect of state ($\chi^2_1 = 56.26$, $p < 0.001$), but did not include a main effect of age or group, and the slopes of the P(stop) and RT relationship did not differ across states or between groups ($ps > 0.145$). In line with recent suggested guidelines for the SST [44], we confirmed that no subjects had an average RT on failed stop trials that was greater than their average RT on go trials. There were no differences by group, state, or group × state interactions for SSRT, post-stop-error go RT (i.e., "post-error slowing"), or stop error RT ($ps > 0.082$; see Supplement).

### fMRI results

*Bayesian expectancy violations (prediction errors).* A significant and large group × state interaction for UPE in the left dorsal caudate indicated that whereas healthy controls showed activation negatively associated with UPE in the fasted state and positively associated with UPE in the fed state, the remitted bulimia nervosa group showed activation positively associated with UPE in the fasted state, and a blunted UPE response in the fed state ($R^2_\beta = 0.22$, 95% CI = 0.09, 0.38; Fig. 3A). In other clusters showing group × state interactions for UPE and in those showing interactions for SPE, prediction-error activation patterns were inconsistent with P(stop)-modulated stop- and go-trial activation patterns (see Supplement).

*Bayesian prediction of inhibitory demand (P(stop)).* There were no group × state interactions for brain activation associated with P(stop) when controlling for variance correlated with trial type or accuracy.

### Associations with clinical variables

Lower fed-state activation tracking UPE in the left dorsal caudate was associated with more frequent past binge eating ($z = -2.998$, $p = 0.003$) and self-induced vomiting ($z = 3.350$, $p < 0.001$; Fig. 3B) in women remitted from bulimia nervosa. Fasted-state caudate signal was unrelated to these past bulimic symptoms ($ps > 0.470$). No associations with duration of illness or remission survived correction for multiple comparisons. Results of exploratory sensitivity analyses are presented in the Supplement.

## DISCUSSION

Little is known about the pathophysiology of the episodic, out-of-control overeating and purging that characterize bulimia nervosa. This investigation tested the hypothesis that metabolic state (being fed or fasted) abnormally affects the computational process of inhibitory control in individuals who have had this eating disorder. Specifically, we used "Bayesian surprise" to examine how women with past bulimia nervosa and healthy comparison women update their internal model to prepare to inhibit behavior when they are in fasted and fed states. A group x state interaction suggested that women remitted from bulimia nervosa showed attenuated activation related to "Bayesian surprise" in the left dorsal caudate in the fed state. This activation tracking Bayesian surprise after eating was lowest among women with more frequent past binge eating and self-induced vomiting. Research in symptomatic individuals is needed. However, these findings support the hypothesis that metabolic state-dependent alterations in inhibitory-control processing contribute to the clinical characteristics of bulimia nervosa. Moreover, they suggest that the integration of computational approaches and neuroimaging can identify brain-based processes that could be targets for future interventions or markers for disease severity in bulimia nervosa.

The dorsal caudate has been associated previously with stimulus-outcome learning, value-based decision-making, prediction and processing of performance feedback, and tracking expectation violations [45, 46]. Via anatomical connections to the prefrontal cortex [47], the dorsal caudate may play a central role in cognitive control by updating control-related predictions [48, 49]. In line with our left-lateralized findings, left dorsal caudate connections with the ventral and dorsolateral prefrontal cortices have been specifically implicated in the implementation of proactive inhibition [50]. Prior research has documented reduced dorsal caudate activation during behavioral inhibition in bulimia nervosa [7, 8]. We detected a state-dependent group difference in dorsal caudate activation tracking the magnitude of control-related expectancy violations (Bayesian UPE). However, we did not detect state-dependent group differences in signals tracking whether the need to stop was unexpectedly present vs. unexpectedly absent (Bayesian SPE), or in activation associated with the expected likelihood that inhibition would be required for each trial (P(stop)). This UPE-specific finding suggests that bulimia

nervosa is linked to altered tracking of the overall discrepancy between model-based expectations about control and experience, rather than an overall failure to predict stop events, altered tracking of expectancy violations that are specific to particular actions, or altered belief updating (see [30]).

Prior work has shown that attenuated neural activation for control-related Bayesian prediction errors in the caudate and lateral prefrontal and anterior cingulate cortices predict the onset of and relapse to problematic substance use [23–25]. Similar to findings in the left dorsal caudate in individuals with methamphetamine dependence [26], the current results suggest that women with a history of bulimia nervosa show reduced left dorsal caudate activation for unexpected demands in the realm of inhibitory control. However, in remitted bulimia nervosa, this alteration was observed specifically after the state change introduced by eating. Our previous research demonstrated that the neural response to taste in reward-related regions is abnormally high after eating in women remitted from bulimia nervosa [18]. We speculate that abnormally low activation tracking of the degree of control expectancy violation may leave these elevated reward signals and high urges to continue eating unchecked. Specifically, because the left caudate plays a central role in inhibition, and UPE signaling is thought to prepare individuals to adaptively adjust behavior, blunted caudate activation for UPE after eating starts may make it difficult for individuals with bulimia nervosa to adjust their control-related strategy to stop eating. It could also promote subsequent out-of-control behaviors like self-induced vomiting. Consistent with this possibility, lower fed-state activation associated with UPE was most pronounced in women with the most frequent past binge eating and purging.

The only other research to date to apply model-based fMRI to bulimia nervosa found attenuated reward prediction-error signals for sucrose tastes in the insula and ventral striatum that were similarly associated with more frequent weekly binge-purge episodes [12]. Therefore, blunted striatal tracking of information about the discrepancy between expectations and experience in both reward and control domains may promote bulimic symptoms.

In addition, our results suggest that metabolic state changes could have opposite effects on this component of inhibitory-control processing in women remitted from bulimia nervosa and healthy controls. We speculate that eating may facilitate adaptive adjustment to inhibit behavior in healthy controls [30], but may have the inverse effect in women remitted from bulimia nervosa. Studies integrating the current methods with more complex tasks, within-task self-report assessments, and neurotransmitter and neuroendocrine measures could help to delineate the underlying mechanisms and consequences of aberrant metabolic state influences in bulimia nervosa. For example, studies incorporating pupillometry, self-reported P(stop) estimates, and confidence ratings for those estimates during SSTs could help to determine whether reduced neural tracking of the predictive accuracy of P(stop) increases individuals' subjective uncertainty about their internal model of how much control is needed after eating.

With regard to behavioral performance, it is possible that the current version of the SST may have failed to detect true group x state effects or group differences. However, some studies find no inhibitory control deficits in symptomatic bulimia nervosa [6, 7]. In particular, two out of the three studies using the SST to study bulimia nervosa to date have found no differences from healthy controls in behavioral accuracy [51]. In addition, the remitted status of our bulimia nervosa group may have contributed to their intact behavioral performance. As impairments in response inhibition are more pronounced on tasks using disorder-relevant stimuli in binge-type eating disorders [51], future studies should investigate the effects of metabolic state on SSTs with food-specific images. Nevertheless, the detected group-by-state interaction in activation associated with Bayesian UPEs and the observed link between altered UPE activation and past symptom severity highlight the utility of model-based neuroimaging in isolating a more precise component of the inhibition process that may go awry in bulimia nervosa.

This study is the first to examine the effects of fasting and eating on inhibitory control-related processes in eating disorders, and the first to apply model-based fMRI to the study of control in eating disorders. Pre-scan nutritional status was monitored and manipulated in a standardized environment, and the tightly controlled repeated-measures design is a major methodological strength. Moreover, our use of the same task and modeling approach as prior studies permits direct comparison of our findings to those in addiction [22–26].

Despite these strengths, study limitations highlight additional directions for future work. The sample size was relatively modest and included only adult females and individuals who had the purging subtype of bulimia nervosa, limiting the generalizability of our findings. Results of sensitivity analyses suggest that our findings were not better accounted for by anxiety or depressive symptoms or by past comorbidities. However, future research should directly compare individuals with bulimia nervosa to those with substance use, mood, or anxiety disorders in various states to test whether reduced striatal activation for control-related surprises represents a transdiagnostic alteration. Finally, because our participants were remitted from bulimia nervosa, we cannot determine whether we detected a trait-level (but state-specific) disturbance, a scar of the illness, or a contributor to recovery. Activation tracking UPE in the fed state was unrelated to duration of illness or to duration of remission, and Altered neural activation associated with inhibition has been previously detected before bulimic symptom onset [52] and early in the course of bulimia nervosa [9]. In our sample, activation tracking UPE in the fed state was unrelated to duration of illness or to duration of remission, perhaps supporting the hypothesis that the observed metabolic state-specific findings represent a premorbid abnormality. If so, changes in other control-related abilities not studied here (e.g., planning) may have been required to compensate for fed-state UPE alterations and support the normalized SST performance and normalized eating behavior observed in our remitted sample. In addition, past dietary restraint and restriction were not measured, limiting our ability to interpret fasted-state activation patterns. An ongoing investigation of the influences of fasting and eating on the neural computations of inhibitory control in an actively symptomatic sample (R01MH132786) aims to build on the current findings to more precisely link state-specific changes in the neural computations underlying cognitive control to the severity and the frequency of current binge eating, purging, and dietary restriction in bulimia nervosa.

## Conclusions and implications for intervention

After receiving first-line interventions, most patients with bulimia nervosa remain symptomatic [53], and a limited understanding of the neural mechanisms underlying bulimic symptoms has thwarted treatment development efforts. Our results are the first to suggest that bulimia nervosa is associated with an altered striatal response for control-related surprises that is modulated by metabolic state. The cognitive-behavioral model of bulimia nervosa [54] proposes that restriction and fasting promote binge eating, and self-control depletion models [55] of bulimia nervosa suggest that restriction drives binge eating specifically by decreasing inhibitory control. As a result, current treatments for bulimia nervosa focus first on eliminating periods of restriction. However, if a blunted signal for control-related surprises makes inhibition more difficult after eating in currently symptomatic individuals, treatments that focus on the fed state may prove fruitful. For instance, interventions that can enhance control

(e.g., non-invasive brain stimulation, behavioral skills practice [56, 57]) may be most effective if delivered during or immediately after food consumption.

## CODE AVAILABILITY

The code used for modeling and analysis is available upon request.

## REFERENCES

1. Udo T, Grilo CM. Prevalence and Correlates of DSM-5–Defined Eating Disorders in a Nationally Representative Sample of U.S. Adults. Biol Psychiatry. 2018;84:345–54.
2. Treasure J, Claudino AM, Zucker N. Eating disorders. Lancet. 2010;375:583–93.
3. Tith RM, Paradis G, Potter BJ, Low N, Healy-Profitós J, He S, et al. Association of Bulimia Nervosa With Long-term Risk of Cardiovascular Disease and Mortality Among Women. JAMA Psychiatry. 2020;77:44–51.
4. Arcelus J, Mitchell A, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. Arch Gen Psychiatry. 2011;68:724–31.
5. Higgs S, Spetter MS, Thomas JM, Rotshtein P, Lee M, Hallschmid M, et al. Interactions between metabolic, reward and cognitive processes in appetite control: Implications for novel weight management therapies. J Psychopharmacol (Oxf, Engl). 2017;31:1460–74.
6. Wu M, Hartmann M, Skunde M, Herzog W, Friederich H. Inhibitory control in bulimic-type eating disorders: a systematic review and meta-analysis. PLoS ONE. 2013;8:e83412.
7. Skunde M, Walther S, Simon J, Wu M, Bendszus M, Herzog W, et al. Neural signature of behavioural inhibition in women with bulimia nervosa. J Psychiatr Neurosci. 2016;41:E69–E78.
8. Marsh R, Steinglass JE, Gerber AJ, O'Leary KG, Walsh BT, Peterson BS. Deficient activity in the neural systems that mediate self-regulatory control in bulimia nervosa. Arch Gen Psychiatry. 2009;66:1–13.
9. Marsh R, Horga G, Wang Z, Wang P, Klahr KW, Berner LA, et al. An fMRI study of self-regulatory control and conflict resolution in adolescents with bulimia nervosa. Am J Psychiatry. 2011;168:1210–20.
10. Shenoy P, Yu AJ. Rational Decision-Making in Inhibitory Control. Front Hum Neurosci. 2011;5:48.
11. Jiang J, Heller K, Egner T. Bayesian modeling of flexible cognitive control. Neurosci Biobehav Rev. 2014;46:30–43.
12. Frank GKW, Reynolds JR, Shott ME, O'Reilly RC. Altered temporal difference learning in bulimia nervosa. Biol Psychiatry. 2011;70:728–35.
13. Loeber S, Grosshans M, Herpertz S, Kiefer F, Herpertz SC. Hunger modulates behavioral disinhibition and attention allocation to food-associated cues in normal-weight controls. Appetite 2013;71:32–9.
14. Bartholdy S, Cheng J, Schmidt U, Campbell IC, O'Daly OG. Task-Based and Questionnaire Measures of Inhibitory Control Are Differentially Affected by Acute Food Restriction and by Motivationally Salient Food Stimuli in Healthy Adults. Front Psychol. 2016;7:1303.
15. Hollmann M, Hellrung L, Pleger B, Schlogl H, Kabisch S, Stumvoll M, et al. Neural correlates of the volitional regulation of the desire for food. Int J Obes. 2012;36:648–55.
16. Thomas JM, Higgs S, Dourish CT, Hansen PC, Harmer CJ, McCabe C. Satiation attenuates BOLD activity in brain regions involved in reward and increases activity in dorsolateral prefrontal cortex: an fMRI study in healthy volunteers. Am J Clin Nutr. 2015;101:697–704.
17. He Q, Huang X, Zhang S, Turel O, Ma L, Bechara A. Dynamic causal modeling of insular, striatal, and prefrontal cortex activities during a food-specific Go/NoGo task. Biol Psychiatry: Cogn Neurosci Neuroimaging. 2019;4:1080–9.
18. Ely AV, Wierenga CE, Bischoff-Grethe A, Bailer UF, Berner LA, Fudge JL, et al. Response in taste circuitry is not modulated by hunger and satiety in women remitted from bulimia nervosa. J Abnorm Psychol. 2017;126:519–30.
19. Berner LA, Simmons AN, Wierenga CE, Bischoff-Grethe A, Paulus MP, Bailer UF, et al. Altered anticipation and processing of aversive interoceptive experience among women remitted from bulimia nervosa. Neuropsychopharmacology 2019;44:1265–73.
20. Wierenga CE, Bischoff-Grethe A, Berner LA, Simmons AN, Bailer U, Paulus MP, et al. Increased anticipatory brain response to pleasant touch in women remitted from bulimia nervosa. Transl Psychiatry. 2020;10:236.
21. Wagner A, Simmons A, Oberndorfer T, Frank G, McCurdy-McKinnon D, Fudge J, et al. Altered sensitization patterns to sweet food stimuli in patients recovered from anorexia and bulimia nervosa. Psychiatr Res. 2015;234:305–13.
22. Ide JS, Shenoy P, Yu AJ, Li C-SR. Bayesian Prediction and Evaluation in the Anterior Cingulate Cortex. J Neurosci. 2013;33:2039.
23. Harlé KM, Shenoy P, Stewart JL, Tapert SF, Yu AJ, Paulus MP. Altered Neural Processing of the Need to Stop in Young Adults at Risk for Stimulant Dependence. J Neurosci. 2014;34:4567–80.
24. Harle KM, Stewart JL, Zhang S, Tapert SF, Yu AJ, Paulus MP. Bayesian neural adjustment of inhibitory control predicts emergence of problem stimulant use. Brain 2015;138:3413–26.
25. Harle KM, Yu AJ, Paulus MP. Bayesian computational markers of relapse in methamphetamine dependence. Neuroimage Clin. 2019;22:101794.
26. Harlé KM, Zhang S, Ma N, Yu AJ, Paulus MP. Reduced Neural Recruitment for Bayesian Adjustment of Inhibitory Control in Methamphetamine Dependence. Biol Psychiatry: Cogn Neurosci Neuroimaging. 2016;1:448–59.
27. Ide JS, Hu S, Zhang S, Yu AJ, Li C-SR. Impaired Bayesian learning for cognitive control in cocaine dependence. Drug Alcohol Depend. 2015;151:220–7.
28. Le TM, Potvin S, Zhornitsky S, Li C-SR. Distinct patterns of prefrontal cortical disengagement during inhibitory control in addiction: A meta-analysis based on population characteristics. Neurosci Biobehav Rev. 2021;127:255–69.
29. Kaye WH, Wierenga CE, Bailer UF, Simmons AN, Wagner A, Bischoff-Grethe A. Does a shared neurobiology for foods and drugs of abuse contribute to extremes of food ingestion in anorexia and bulimia nervosa? Biol psychiatry. 2013;73:836–42.
30. Yu AJ, Cohen JD. Sequential effects: superstition or rational behavior? Adv Neural Inf Process Syst. 2008;21:1873–80.
31. Rouhani N, Niv Y. Signed and unsigned reward prediction errors dynamically enhance learning and memory. eLife 2021;10:e61077.
32. Valentin VV, O'Doherty JP. Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. J Neurophysiol. 2009;102:3384–91.
33. Colas JT, Pauli WM, Larsen T, Tyszka JM, O'Doherty JP. Distinct prediction errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fMRI. PLOS Comput Biol. 2017;13:e1005810.
34. Asaad WF, Eskandar EN. Encoding of Both Positive and Negative Reward Prediction Errors by Neurons of the Primate Lateral Prefrontal Cortex and Caudate Nucleus. J Neurosci. 2011;31:17772.
35. Jiang J, Beck J, Heller K, Egner T. An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. Nat Commun. 2015;6:8165.
36. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: Fouth Edition, Text Revision (DSM-IV-TR). Washington, DC: American Psychiatric Association; 2000.
37. Wagner A, Aizenstein H, Frank G, Figurski J, May C, Putnam K, et al. Taste challenge reveals altered insula response in anorexia nervosa after recovery - an fMRI study. Biol Psychiatry. 2006;59:202S.
38. Wierenga C, Bischoff-Grethe A, Melrose A, Grenesko-Stevens E, Irvine Z, Wagner A, et al. Altered BOLD response during inhibitory and error processing in adolescents with anorexi anervosa. PLoS ONE. 2014;9:392017.
39. Oberndorfer T, Kaye W, Simmons A, Strigo I, Matthews S. Demand-specific alteration of medial prefrontal cortex response during an ihhibition task in recovered anorexic women. Int J Eat Disord. 2011;44:1–8.
40. Matthews S, Simmons A, Arce E, Paulus M. Dissociation of inhibition from error processing using a parametric inhibitory task during functional magnetic resonance imaging. Neuroreport 2005;16:755–60.
41. Harlé KM, Spadoni AD, Norman SB, Simmons AN. Neurocomputational Changes in Inhibitory Control Associated With Prolonged Exposure Therapy. J Trauma Stress. 2020;33:500–10.
42. Howlett JR, Harlé KM, Paulus MP. Inhibitory failures in cocaine use disorder: Not paying attention when there is a need to be cautious. Drug Alcohol Depend. 2021;226:108833.
43. Eklund A, Nichols T, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci USA. 2016;113:7900–5.
44. Verbruggen F, Aron AR, Band GPH, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. eLife 2019;8:e46323.
45. Elliott R, Frith CD, Dolan RJ. Differential neural response to positive and negative feedback in planning and guessing tasks. Neuropsychologia 1997;35:1395–404.
46. D'Astolfo L, Rief W. Learning about Expectation Violation from Prediction Error Paradigms – A Meta-Analysis on Brain Processes Following a Prediction Error. Front Psychol. 2017;8:1253.
47. Draganski B, Kherif F, Kloppel S, Cook PA, Alexander DC, Parker GJ, et al. Evidence for segregated and integrative connectivity patterns in the human Basal Ganglia. J Neurosci: Off J Soc Neurosci. 2008;28:7143–52.
48. Frank MJ. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci. 2005;17:51–72.

49. Frank MJ, Badre D. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. Cereb Cortex. 2012;22:509–26.
50. Zhang F, Iwaki S. Common Neural Network for Different Functions: An Investigation of Proactive and Reactive Inhibition. Front Behav Neurosci. 2019;13:124.
51. Wu M, Giel KE, Skunde M, Schag K, Rudofsky G, de Zwaan M, et al. Inhibitory control and decision making under risk in bulimia nervosa and binge-eating disorder. Int J Eat Disord. 2013;46:721–8.
52. Bartholdy S, O'Daly OG, Campbell IC, Banaschewski T, Barker G, Bokde ALW, et al. Neural Correlates of Failed Inhibitory Control as an Early Marker of Disordered Eating in Adolescents. Biol Psychiatry. 2019;85:956–65.
53. Linardon J, Wade TD. How many individuals achieve symptom abstinence following psychological treatments for bulimia nervosa? A meta-analytic review. Int J Eat Disord. 2018;51:287–94.
54. Fairburn CG. Cognitive Behavior Therapy and Eating Disorders. New York: Guilford Press; 2008.
55. Loth KA, Goldschmidt AB, Wonderlich SA, Lavender JM, Neumark-Sztainer D, Vohs KD. Could the resource depletion model of self-control help the field to better understand momentary processes that lead to binge eating? Int J Eat Disord. 2016;49:998–1001.
56. Kekic M, McClelland J, Bartholdy S, Boysen E, Musiat P, Dalton B, et al. Single-Session Transcranial Direct Current Stimulation Temporarily Improves Symptoms, Mood, and Self-Regulatory Control in Bulimia Nervosa: A Randomised Controlled Trial. PLOS ONE. 2017;12:e0167606.
57. Preuss H, Pinnow M, Schnicker K, Legenbauer T. Improving inhibitory control abilities (ImpulsE)—A promising approach to treat impulsive eating? Eur Eat Disord Rev. 2017;25:533–43.

## ACKNOWLEDGEMENTS

Collection of these data was supported by grants from the NIH (R01MH042984-17A1 and R01MH042984-18S1) and the Price Foundation (PI: WHK). Data analysis and preparation of this paper were supported by a career development award from the NIH (K23MH118418; PI: LAB) and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (PI: LAB). Portions of these findings were presented by LAB at the December 2019 meeting of the American College of Neuropsychopharmacology in Orlando, FL, as well as the October 2020 meeting of the Eating Disorders Research Society and the November 2020 meeting of the Association for Behavioral and Cognitive Therapies (both held remotely due to COVID-19). The authors thank Miki Carrico for assistance with neuroimaging data collection and preprocessing. We would also like to thank Daria Orlowska for assistance with participant screening and data collection. Finally, we thank the women who participated in this study for their time.

## AUTHOR CONTRIBUTIONS

LAB: Conceptualization, Methodology, Formal Analysis, Visualization, Writing - Original Draft Preparation, Review, and Editing; KMH, ANS, AY, MPP: Methodology, Writing - Review and Editing; ABG: Methodology, Data Curation, Writing - Review and Editing; CEW: Project Administration, Supervision, Writing - Review and Editing; UFB: Investigation, Supervision; WHK: Funding Acquisition, Conceptualization, Methodology, Resources, Supervision, Writing - Review and Editing. All authors reviewed and approved the final version of the paper.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41380-023-02063-6.

**Correspondence** and requests for materials should be addressed to Laura A. Berner.