International Journal of
EATING DISORDERS

When Food Becomes a Distraction: The Impact of Food-Related Information on Learning in Anorexia Nervosa

Journal:	*International Journal of Eating Disorders*
Manuscript ID	IJED-23-0307
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	30-May-2023

Complete List of Authors:	Caudek, Corrado; University of Florence, NEUROFARBA Colpizzi, Ilaria ; University of Florence, Department of Health Sciences Guidi, Lisa; Università degli Studi di Firenze, NEUROFARBA Danti, Sara; Università degli Studi di Firenze, NEUROFARBA Lucchesi, Silvia; Università degli Studi di Firenze, NEUROFARBA Giusti, Elisa; Università degli Studi di Firenze, NEUROFARBA Di Meglio, Martina; Università degli Studi di Firenze, NEUROFARBA Ballardini, Donatella; Università degli Studi di Firenze, NEUROFARBA Mazzoni, Chiara; Università degli Studi di Firenze, NEUROFARBA Pieraccioli, Caterina; Università degli Studi di Firenze, NEUROFARBA Schumann, Romana; Università degli Studi di Firenze, NEUROFARBA Sica, Claudio; Università degli Studi di Firenze, Department of Health Sciences

Keywords:	anorexia nervosa, reinforcement learning, domain-specificity, probabilistic reversal learning

When Food Becomes a Distraction: The Impact of Food-Related Information on Learning in Anorexia Nervosa

Corrado Caudek[1], Ilaria Colpizzi[2], Lisa Guidi[3], Sara Danti[3], Silvia Lucchesi[3], Elisa Giusti[3], Martina Di Meglio[3], Donatella Ballardini[4], Chiara Mazzoni[4], Caterina Pieraccioli[5], Romana Schumann[4], & Claudio Sica[2]

[1] Department NEUROFARBA, University of Florence, Italy

[2] Heath Sciences Department, University of Florence, Italy

[3] Istituto Specchidacqua, Frida Onlus Montecatini, Italy

[4] Centro Gruber, Bologna, Italy

[5] Neomesia, Villa dei Pini, Firenze, Italy

Author note

The authors made the following contributions. Corrado Caudek: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing, Statistical analyses, Supervision; Ilaria Colpizzi: Conceptualization, Writing - Review & Editing, Data Collection, Statistical analyses; Lisa Guidi: Conceptualization, Data Collection, Writing - Review & Editing; Sara Danti: Conceptualization, Data Collection, Writing - Review & Editing; Silvia Lucchesi: Conceptualization, Data Collection, Writing - Review & Editing; Elisa Giusti: Conceptualization, Data Collection, Writing - Review & Editing; Martina Di Meglio: Conceptualization, Data Collection, Writing - Review & Editing; Donatella Ballardini: Conceptualization, Data Collection, Writing - Review & Editing; Chiara Mazzoni: Conceptualization, Data Collection, Writing - Review & Editing; Caterina Pieraccioli: Conceptualization, Data Collection, Writing - Review & Editing; Romana Schumann: Conceptualization, Data Collection, Writing - Review & Editing; Claudio Sica: Conceptualization, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Corrado Caudek, via di San Salvi n. 12, Padiglione 26, Firenze, 50139, Italy. E-mail: corrado.caudek@unifi.it

Word count: 4423

Disclosure statement. The authors declare that they have no conflict of interest.

Abstract

Objective: This study aimed to investigate the performance of individuals with anorexia nervosa (AN) in Reinforcement Learning (RL) tasks under different contexts, specifically examining if RL abilities are unaffected in neutral contexts but significantly impaired in food-related contexts. **Method**: RL performance was assessed using a Probabilistic Reversal Learning (PRL) task, with participants exposed to outcome-irrelevant food-related information or neutral information. A clinical sample (N = 49) was compared to a control group (N = 229). **Results**: AN individuals demonstrated lower learning rates for food-related decisions, while their performance on neutral decisions was comparable to participants with Bulimia Nervosa, Healthy Controls (HCs), and HCs at risk of eating disorders. Additionally, only AN patients exhibited reduced learning rates for outcome-irrelevant food-related decisions in reward-based learning, as opposed to food-unrelated decisions. **Discussion**: Impaired RL task performance in individuals with AN may be attributed to external factors rather than compromised learning mechanisms. These findings indicate that AN may significantly impact the cognitive processing of food-related information, even when AN patients do not show learning rate disadvantages compared to HCs in decision-making involving food-unrelated information. This study provides valuable insights into the reinforcement learning processes of individuals with AN and emphasizes the need to consider the influence of food-related information on cognitive functioning in this patient population. The findings have potential implications for the development of interventions targeting decision-making processes in individuals with AN.

Public significance statement: Impaired RL task performance in individuals with AN is primarily influenced by external factors, rather than compromised learning mechanisms. Our findings challenge the notion that general associative learning dysfunction is the underlying issue, suggesting that interventions targeting such dysfunction may be ineffective. Instead, interventions should prioritize addressing factors that hinder unimpaired associative learning abilities, such as food-irrelevant information, symptom-related contexts, or long-term goals.

Keywords: anorexia nervosa, bulimia nervosa, reinforcement learning, domain-specificity

When Food Becomes a Distraction: The Impact of Food-Related Information on Learning in Anorexia Nervosa

Introduction

Eating disorders are severe and notoriously difficult to treat psychiatric conditions, which have prompted the need for a deeper understanding of their mechanisms of development and maintenance (Fairburn, 2008). Recent research has suggested a potential role for executive processes in the pathophysiology of eating disorders. Among these, impairments in cognitive inflexibility (Wu et al., 2014), decision-making (Guillaume et al., 2015), and inhibitory control (Bartholdy et al., 2016) have been linked to eating disorders. Cognitive inflexibility has been the most extensively studied, particularly through the use of a Reinforcement Learning (RL) paradigm. Although the hypothesis of maladaptive associative learning appears theoretically compelling and holds potential for treatment, the evidence supporting it remains inconsistent (Caudek et al., 2021). This inconsistency in the literature has prompted a recent proposal suggesting an alternative explanation for the impaired performance that, under some conditions, is observed in individuals with Anorexia Nervosa (AN) and other eating disorders (EDs) during reinforcement learning (RL) tasks. Caudek et al. (2021), Haynos et al. (2022a), and Haynos et al. (2022b) propose that the compromised performance in RL tasks may not be attributed to compromised learning mechanisms within individuals, but rather to external factors that disrupt the learning process.

Recent research has indeed demonstrated that features unrelated to outcomes can affect RL and decision-making in the general population (*e.g.,* Ben-Artzi et al., 2022; Caudek et al., 2021). Consistent with these results, the present study will test the hypothesis that the impaired performance of individuals with Anorexia Nervosa (AN) in reinforcement learning (RL) tasks can be attributed to external factors that disrupt the learning process, rather than intrinsic deficiencies in their learning mechanisms. Such external factors could comprise spatial-motor associations, which have been found to impact reinforcement learning in the general population (Shahar et al., 2019), as well as disease-related factors, such as long-term goals and personality traits (Haynos et al., 2020). Specifically, we propose that individuals diagnosed with AN may experience interference in their decision-making process while making choices between a food and non-food item in a Probabilistic Reversal Learning (PRL) task, where the image's content is inconsequential to its reward value. Therefore, we suggest that AN patients' long-term goals linked to weight control can impact their decision-making when food-related information is present, even if it is not pertinent to the outcome, thereby hindering their decision-making ability (see also Haynos et al., 2022a).

In technical terms, we anticipate that individuals with AN would exhibit a conservative learning behavior by updating their expectations more slowly in response to feedback from their previous choices. This hypothesis suggests that individuals with AN may possess normal cognitive decision-making skills, but external factors, such as long-term goals related to weight control, can interfere with their decision-making processes, resulting in impaired performance in RL tasks (that is, it can make them conservative learners).

According to the context-dependent conservative learning hypothesis, we can make several predictions regarding the impact of outcome-irrelevant features on PRL performance. Firstly, based on previous research on attention and cognitive control for food-related versus food-unrelated information (Caudek et al., 2021; Schiff et al., 2021), we anticipate that both individuals with eating disorders and Healthy Controls (HCs) will demonstrate a more cautious approach to food-related information than to neutral food-unrelated information. In terms of computational modeling, this cautious approach will be indicated by an increase in the threshold for decision commitment, meaning that a greater amount of evidence will be required before a decision is made in response to food-related information relative to neutral food-unrelated information. Secondly, individuals with AN are expected to exhibit reduced learning rates from positive or negative feedback compared to healthy controls when outcome-irrelevant food-related information is present. This effect is specific to food-related information. Thirdly, individuals with AN are anticipated to require a greater amount of evidence before committing to a decision as compared to healthy controls, but only when outcome-irrelevant food-related information is present. Fourthly, individuals with AN are expected to exhibit a reduced learning rate in the presence of food-related information, as compared to neutral food-unrelated information, whereas such a difference will not be observed among healthy controls. Finally, we expect context-dependent conservative learning to be unique to individuals with AN, not only when compared to HCs but also to those at risk of developing eating disorders (RI), and in contrast to individuals with Bulimia Nervosa (BN), who may be more impulsive (Howard et al., 2020).

This investigation has the potential to shed light on how individuals with eating disorders are influenced by external factors, including long-term goals and personality traits, which may significantly impact their decision-making processes, independent of the immediate outcomes of their choices. The identification of these factors and their effects on decision-making in this population could have considerable implications for understanding the underlying mechanisms of eating disorders and developing more effective treatment interventions (*e.g.*, Haynos et al., 2020).

Methods

The study, which adhered to the Declaration of Helsinki, was approved by the University of Florence's Ethical Committee (Prot. n. 0178082). All eligible participants provided informed consent and agreed to participate.

Participants

The study recruited a total of 40 patients diagnosed with AN, 13 patients with BN, 213 HCs, and 33 healthy individuals at risk of developing eating disorders. The patients received outpatient treatment at three different facilities in Italy, namely the Specchidacqua Institute in Montecatini (Pisa), the Villa dei Pini Institute in Firenze, and the Gruber Center, Outpatient Clinic in Bologna. Specialized psychiatrists and psychologists at these institutes conducted psychiatric evaluations based on the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) criteria. Patients with AN and BN were assessed approximately 6 months ($\pm$ 1 month) after initiating treatment for eating disorders at one of the participating facilities. The present study assessed the eligibility criteria for participants via clinical interviews conducted by trained psychologists. Participants were required to exhibit proficient command over both spoken and written Italian language, as well as report cognitive function within the normal range, as assessed by the Raven's Standard Progressive Matrices test (Raven et al., 2000). Neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis were the exclusion criteria for participation in this study. Comorbid diagnoses were ascertained through psychiatric evaluations conducted over a period of no less than 6 months, while the absence of comorbidities was assessed utilizing identical methodologies within an equivalent time frame.

The HC group was recruited through social media or university advertisements. Moreover, individuals at risk of developing eating disorders were enrolled from the University of Florence community using the Eating Attitudes Test-26 (EAT-26; Garner et al., 1982) screening tool. Participants who scored higher than 20 on the EAT-26 and did not report any current treatment for eating disorders were categorized as "at-risk" individuals for the purpose of this study (Dotti & Lazzari, 1998). The eligibility criteria for both the HCs and the individuals at risk of developing eating disorders were evaluated through psychologist interviews. Exclusion criteria encompassed abnormal body mass index (BMI) values (determined in the laboratory), neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis.

The majority of the participants in this study were of Caucasian ethnicity (97.7%), followed by Asian-Italian (1.7%) and African-Italian (0.6%). Furthermore, all participants were right-handed and were kept blind to the study's objectives.

Procedure

During the initial session, participants underwent a clinical interview to determine their eligibility for the study. Those who met the criteria proceeded to anthropometric measurements and were asked to complete the psychometric scales listed below. In a subsequent session, participants completed the PRL task and were subsequently provided with a debriefing.

We compared the characteristics of the clinical sample with the controls by administering the following scales: the EAT-26, the Body Shape Questionnaire-14 (BSQ-14; Dowson & Henderson, 2001), the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998), the Depression Anxiety Stress Scale-21 (DASS-21; Lovibond & Lovibond, 1995), the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965), the Multidimensional Perfectionism Scale (MPS-F; Frost et al., 1990), and the Raven's Standard Progressive Matrices (Raven et al., 2000). The results of these statistical analyses are provided in the Supplementary Information (SI).

The PRL task consisted of two blocks, each containing 160 trials (see Figure 1). In each trial of the PRL task, participants were required to select one image within a 3-second time limit. After 0.5 s, a euro coin image was presented as a reward for a correct response, while a strike-through image of a euro coin was used as a punishment for an incorrect response. Feedback was provided for 2 seconds after each trial. A random inter-trial interval (blank screen) between 0.5 and 3 s was used. The PRL task consisted of four epochs, with each epoch containing 40 trials, and the same image category was considered correct throughout each epoch. Feedback was probabilistic, with the correct image being rewarded in 70% of cases per epoch, while negative feedback was provided in the remaining 30% of trials. Both blocks of the task included three rule changes in the form of a reversal phase. Participants were informed that stimulus-reward contingencies would change, but not the specifics of how or when this would occur. The participants' objective was to maximize their earnings, which were displayed at the end of each block (for further information, see the SI).

Data analysis

The data were analyzed using Bayesian statistical methods. Credible effects were determined by examining 95% credible intervals or by assessing whether 97.5% of posterior samples fell above or below 0 when computing the proportion of posterior in the direction of the effect.

In order to assess domain-specific biases in learning, we employed the Reinforcement Learning Drift Diffusion Model (RLDDM; Pedersen et al., 2017; Pedersen & Frank, 2020) to examine the two-choice decision-making process over time in the PRL task. Cognitive modeling analysis provides a robust approach to deconstructing decision-making task performance into its constituent processes, thus revealing underlying mechanisms that may not be evident from the overall task outcome.

Transparency and Openness

We report all data exclusion criteria and how the sample size was determined. All measures used in this study are reported. Data, and analysis code are available upon request to the corresponding author. Data were analyzed using Python and R version 4.3. This study was not preregistered.

Results

Demographic and Psychopathology Measures

Mean age and Body Mass Index (BMI) for each group of participant were as follows: patients with AN, mean age = 21.18 (*SD* = 2.41), average Body Mass Index (BMI) = 16.88 (*SD* = 1.55); patients with BN, mean age = 20.39 (*SD* = 1.88), average BMI = 30.09 (*SD* = 5.47); HCs, mean age = 19.77 (*SD* = 1.06), average BMI = 21.62 (*SD* = 3.03); healthy individuals at risk of developing eating disorders, mean age = 20.36 (*SD* = 1.44), average BMI = 22.41 (*SD* = 4.79).

Bayesian statistical analysis revealed no credible age differences among the four groups (AN, BN, HC, and RI). AN participants displayed a lower mean BMI than HC participants, while BN participants had a higher mean BMI than HC participants. No noteworthy difference in BMI was observed between HC and RI participants. Furthermore, there is credible evidence that the Rosenberg Self-Esteem Scale scores of all three groups (AN, BN, and RI) are smaller than those of the HC group. We also found credible evidence that individuals with AN, BN, and RI exhibited higher levels of dissatisfaction with their body shape, as measured by the BSQ-14 questionnaire, when compared to the HCs.

Individuals with AN displayed higher stress, anxiety, and depression levels (as measured by the DASS-21) than HCs. Additionally, individuals with AN showed credibly higher levels of social interaction anxiety (as measured by the SIAS) than HCs. All three AN, BN, and RI groups exhibited higher levels of Concerns over mistakes and doubts scores of the MPS scale compared to HCs. Individuals with AN also showed higher levels of Personal standard scores of the MPS scale compared to HCs. Moreover, individuals with AN displayed higher values on all three subscales of the EAT-26 questionnaire relative to HCs. For more detailed information regarding these comparisons, refer to the Supplementary Information (SI).

Sixteen individuals with AN were diagnosed with a comorbid anxiety disorder, 8 with OCD, 1 with social phobia, and 1 with DAP; among the individuals with BN, 4 were diagnosed with mood disorder and 1 with OCD.

Reinforcement learning and drift diffusion modeling

To test the interference of disease-related information on the decision process, we compared several RLDDMs in which we conditioned either none, each or all model's parameters on diagnostic category (group) and image category (food-unrelated and food-related). For each model, we computed the Deviance Information Criterion (DIC), and we selected the model with the best trade-off between the fit quality and model complexity (i.e., the model with the lowest DIC).

The following RLDDM models were examined. Model M1 is a standard RLDDM; starting point was fixed at 0.5 (unbiased priors). Model M2 extends M1 by incorporating separate learning rates for positive and negative reinforcements. In Model M3, the α^+ and α^- parameters are based on the diagnostic group. In Model M4, the α^+ and α^- parameters of M3 are conditioned on both diagnostic group and image category (two neutral images, or one neutral and one symptom-related image). Model M5 expands upon M4 by considering that the a parameter may be influenced by both diagnostic group and image category. Model M6 extends M5 by taking into account the possible influence of diagnostic group and image category on the v parameter. Model M7 builds upon M6 by considering that the t parameter may depend on both diagnostic group and image category. The winning RLDDM, as determined by the lowest deviance information criterion (DIC), is M7 (see Table 1). In this model, the parameters α^+, α^-, a, v, and t are conditioned on both diagnostic groups and the presence of food-related or food-unrelated outcome-irrelevant information.

Convergence was evaluated through the examination of trace and autocorrelation plots, as well as assessment of the Gelman-Rubin statistic for Bayesian model parameters. All model parameters exhibited $\hat{R}$ values below 1.1 (max = 1.071, mean = 1.002), which indicates an absence of convergence issues. Posterior predictive checks were used to evaluate model validity.

To gauge the impact of outcome-irrelevant image category on decision-making, we computed the difference in posterior estimates of the parameters of the RLDDM between the food-unrelated and food-related conditions for each group. To examine Hypothesis H1, we conducted a comparison of decision thresholds (a) for food-related and food-unrelated information across groups. The results of this analysis are presented in Figure 2. Hypothesis H2 was evaluated through a comparison of the learning rates (α^+, α^-) of individuals with AN to those of HCs, for both food-related and food-unrelated outcome-irrelevant information (see Figures 3 and 4). In order to test Hypothesis H3, we compared the decision thresholds (a) of individuals with AN to those of HCs for both food-related and food-unrelated outcome-irrelevant information (see Figures 5. To evaluate Hypothesis H4, we conducted a comparison of the learning rate for rewards (α^+) for food-related and food-unrelated information (α^+ food - related $- \alpha^+$ food - unrelated) across groups (see Figure 6). To examine Hypothesis H5, we compared the RLDDM parameters of the BN and RI groups to those of the HC group. The contrasts between conditions were conducted on the linear predictor scale, with the logit transformation applied to learning rates.

The results presented in the figures provide empirical support for Hypotheses 1, 2, and 3. First, the results indicate that decision thresholds for food-related outcome-irrelevant information are higher than for food-unrelated information, across all diagnostic groups. Second, the learning rates for both positive (α^+) and negative (α^-) feedback are lower for individuals with AN compared to HCs, but only in the presence of food-related information. Third, the decision thresholds are higher for individuals with AN compared to HCs, but only in the presence of food-related information. Fourth, we partially confirmed Hypothesis 4, which examines the discrepancy in learning rates for food-related and food-unrelated information. We observed a difference in the learning rate for rewards (α^+) between the two types of information in individuals with AN. However, we did not find any credible differences in the learning rates for punishments (α^-) between the two types of information, as indicated in the SI. Furthermore, we confirmed Hypothesis 5, as we found no credible differences among the posterior estimates of the RLDDM models for individuals with BN, HCs at risk of developing eating disorders, and HCs (for details on the statistical analyses, see the SI).

Comorbidity

To investigate the potential influence of comorbid conditions on the observed conservative learning behavior displayed by individuals with AN in response to food-related information that is irrelevant to outcome, we employed model M7 on AN patients by grouping them into those with and without comorbidities. Our statistical analysis demonstrated no credible differences in parameters between the two groups (for details, see the SI).

Discussion

All our hypotheses were supported, with some caveats. Firstly, the inclusion of outcome-irrelevant food-related information (in contrast to food-unrelated information) increased the evidence required for decision-making in a PRL task across all participant groups (a parameter in the RLDDM), confirming H1. Secondly, individuals with AN have lower learning rates for food-related decisions (but not neutral decisions) compared to HCs (α^+ and α^- parameters of the RLDDM), supporting H2. Thirdly, individuals with AN exhibit higher decision thresholds than HCs, but only when outcome-irrelevant food-related information is present (a parameter in the RLDDM), validating H3. Fourthly, compared to healthy controls, individuals with AN demonstrate diminished learning rates for outcome-irrelevant food-related decisions in contrast to food-unrelated decisions, specifically in reward-based learning and not in punishment-based learning, confirming H4. While no a priori hypothesis was made regarding the difference in learning rates between rewards and punishments, the findings suggest that the negative impact of outcome-irrelevant information on learning rates is more pronounced in individuals with AN during reward-based learning than in punishment-based learning. Finally, individuals with BN and HCs at risk for eating disorders show comparable results to HCs and dissimilar results from individuals with AN across all comparisons, supporting H5. In conclusion, these findings suggest that individuals with AN exhibit changes in their decision-making processes in the presence of food-related information, regardless of its relevance to the outcome, indicating that AN may fundamentally impact the cognitive processing of food-related information.

The results of this study emphasize the importance of extraneous information that is irrelevant to the outcome in shaping decision-making within the context of reinforcement learning. While all participants displayed a more cautious decision-making style when presented with outcome-irrelevant food information in a probabilistic reinforcement learning task, only individuals with AN exhibited changes in their learning rates in response to such information. These findings support the hypothesis that exposure to food-related cues that are irrelevant to the outcome prompts a shift towards conservative learning in individuals with AN, which is not observed in food-unrelated decision-making contexts.

Our results concerning the comparison between individuals with AN and BN are consistent with previous studies that have reported differences in decision-making between these two categories of patients (*e.g.,* Chan et al., 2014). However, our study also highlights that these differences are more pronounced when considering the processing of information related to the disorder.

We interpret the results of this study as suggesting that learning in individuals with AN may be domain-specific, as defined by Spunt and Adolphs (2017), who suggested that the higher-order beliefs and goals could dynamically regulate the internal operations of a processing module, like RL, through attention and context. The findings of the study support this proposal, as individuals with AN performed similarly to healthy controls in RL tasks involving food-unrelated contexts but exhibited more conservative learning behavior when exposed to food-related information, as indicated by their lower learning rates.

The present results (together with the proposal of Haynos et al., 2020) are at odd with current theories that characterize AN solely in terms of a combination of reduced sensitivity to reward and increased sensitivity to punishment, leading to an imbalance in reward processing. This imbalance is thought to result in decreased interest in food rewards and increased control over food intake, which contributes to the persistence of AN symptoms. Additionally, heightened punishment sensitivity is thought to contribute to AN by promoting avoidance of food and weight gain, which may be perceived as aversive.

Our findings challenge the notion that AN is solely characterized by distorted reward and punishment processing, as this domain-general description fails to account for the nuances in response based on cue characteristics. This argument is supported by Haynos et al. (2020), who presents evidence that suggests a lack of universal deficit in reward and punishment processing among individuals with AN. Instead, Haynos et al. (2020) proposes that individuals with AN exhibit an incorrect interpretation of what constitutes a reward or punishment in different contexts and for various stimuli and decisions. Haynos et al. (2020) suggests that behaviors that are initially perceived as neutral can eventually become associated with either positive or negative reactions, leading them to serve as a form of reward or punishment. For instance, restrictive eating cues, a precursor of AN, can be linked to reward responses in AN, as revealed by ecological momentary assessment (EMA) studies that examined affective patterns in relation to disordered eating. These studies have shown higher positive affect and lower negative affect before, during, and after restrictive eating episodes in AN compared to normal meals (Fitzsimmons-Craft et al., 2015) and subsequent reductions in guilt and increased self-assurance for individuals with AN (Haynos et al., 2017). These findings indicate that restrictive eating is linked to desirable emotional outcomes in AN and, thus, can be understood as rewarding.

Although decreased sensitivity to reward in AN has been documented in some contexts, such as individuals with AN scoring lower on sensation-seeking measures that gauge reactions to immediate novel rewards compared to healthy individuals and those with BN or binge eating disorder (Matton et al., 2015; Rotella et al., 2018), this does not indicate that a reduced sensitivity to reward is evident across all contexts. For instance, the rewarding nature of restrictive eating is not reflected in this reduced sensitivity. The review by Haynos et al. (2020) offers several additional examples of cues, contexts, or decisions that may only be associated with reward or punishment if they are viewed in the context of the ultimate objectives of AN (i.e., thinness). This way of thinking is very much in line with the present results. What the present study adds to this previous theoretical proposal is that previous evidence of domain-specificity of reward and punishment processing in AN have been provided in an indirect form, that is, in terms of the *re-interpretation* of cues and consequences of actions in the context of an overarching long-term goal. In other words, these previous studies have primarily examined the subjective value assigned by AN patients to various experiences, which may be perceived as rewarding or punishing, despite not inherently having these properties. In contrast, the current study, for the first time, investigates the effect of contextual factors on the learning process in which reward and punishment are direct consequences of choices.

From a translational standpoint, it is important to acknowledge that current strategies, such as Cognitive Remediation Therapy (CRT), employ behavioral and cognitive training to augment cognitive flexibility in individuals diagnosed with eating disorders (EDs) (Tchanturia et al., 2010). However, recent studies have suggested that this approach may offer only limited effectiveness (Hagan et al., 2020).

In our study, we found that individuals with AN (at the particular disease stage that we examined) demonstrated a conservative strategy exclusively while learning and processing the consequences of their food choices. Our findings thus suggest that clinical interventions targeting any possible general associative learning dysfunction may not be effective at this stage. Rather, we speculate that interventions should focus on addressing factors that obstruct the performance of unimpaired associative learning abilities, such as food-irrelevant information or symptom-related context, as also suggested by Trapp et al. (2022).

There remain many questions for future research. (1) For example, we used images of a one euro coin or a barred representation of a one euro coin to symbolize rewards and punishments, respectively. But such rewards and punishments are only symbolic and the question remains as to what happens when the rewards and punishments are concrete and not symbolic. (2) Our study only included AN patients who were not in the most severe stage of the illness, as they were recruited from a center for individuals seeking voluntary medical and psychological support. We did not consider AN patients who are hospitalized due to the life-threatening nature of their illness. It is possible that at the later stage of the illness, the associative learning abilities, which were shown to be preserved in the present sample under neutral conditions, may become impaired. (3) Our findings indicate that AN patients displayed no difference in their choice behavior, as assessed by the relative frequency of image selections, when presented with a choice between a food-related and a food-unrelated image (although the inclusion of food-related outcome-irrelevant information had a negative impact on their learning rate). While the relative frequency of image selections failed to detect whether there are differences in the relative "salience" of food-related / food-unrelated images in the PRL task, other measures such as fixation length or number of fixations may provide better insights in this regard. Therefore, future studies could investigate these alternative measures to better capture differences in salience between food-related and food-unrelated images in AN patients. (4) The notion that domain-specificity can be construed as a flexible modulation of a cognitive mechanism across contexts and time (Spunt & Adolphs, 2017), in conjunction with our current observation that individuals with AN exhibit slow learning only in the domain of food-related decision-making, raises the prospect of utilizing the present domain-specificity findings as an evaluative tool. Specifically, a reduction in the effect described in the present study could serve as a marker for the effectiveness of treatment. This promising avenue could be explored in future investigations. (5) In our study, we excluded women under the age of 18. However, this age range is a critical period, as the onset of AN during this stage may have a more profound impact on associative learning, given that cognitive development is ongoing and protective factors are less developed. Future studies should take this into consideration.

References

Bartholdy, S., Dalton, B., O'Daly, O. G., Campbell, I. C., & Schmidt, U. (2016). A systematic review of the relationship between eating, weight and inhibitory control using the stop signal task. *Neuroscience & Biobehavioral Reviews*, *64*, 35–62.

Ben-Artzi, I., Luria, R., & Shahar, N. (2022). Working memory capacity estimates moderate value learning for outcome-irrelevant features. *Scientific Reports*, *12*(1), 1–10.

Caudek, C., Sica, C., Cerea, S., Colpizzi, I., & Stendardi, D. (2021). Susceptibility to eating disorders is associated with cognitive inflexibility in female university students. *Journal of Behavioral and Cognitive Therapy*, *31*(4), 317–328.

Chan, T. W. S., Ahn, W.-Y., Bates, J. E., Busemeyer, J. R., Guillaume, S., Redgrave, G. W., … Courtet, P. (2014). Differential impairments underlying decision making in anorexia nervosa and bulimia nervosa: A cognitive modeling analysis. *International Journal of Eating Disorders*, *47*(2), 157–167.

Dotti, A., & Lazzari, R. (1998). Validation and reliability of the italian EAT-26. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *3*(4), 188–194.

Dowson, J., & Henderson, L. (2001). The validity of a short version of the Body Shape Questionnaire. *Psychiatry Research*, *102*(3), 263–271.

Fairburn, C. G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press.

Fitzsimmons-Craft, E. E., Accurso, E. C., Ciao, A. C., Crosby, R. D., Cao, L., Pisetsky, E. M., Le Grange, D., Peterson, C.B., Crow, S.J., Engel, S.G. Mitchell, J.E., & Wonderlich, S. A.

(2015). Restrictive eating in anorexia nervosa: Examining maintenance and consequences in the natural environment. *International Journal of Eating Disorders*, *48*(7), 923-931.

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, *14*(5), 449–468.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*(4), 871–878.

Geisler, D., Ritschel, F., King, J. A., Bernardoni, F., Seidel, M., Boehm, I., Runge, F., Goschke, T., Roessner, V., Smolka, M. N., & Ehrlich, S. (2017). Increased anterior cingulate cortex response precedes behavioural adaptation in anorexia nervosa. *Scientific Reports*, *7*(1), 1-10.

Guillaume, S., Gorwood, P., Jollant, F., Van den Eynde, F., Courtet, P., & Richard-Devantoy, S. (2015). Impaired decision-making in symptomatic anorexia and bulimia nervosa patients: A meta-analysis. *Psychological Medicine*, *45*(16), 3377–3391.

Hagan, K. E., Christensen, K. A., & Forbush, K. T. (2020). A preliminary systematic review and meta-analysis of randomized-controlled trials of cognitive remediation therapy for anorexia nervosa. *Eating Behaviors*, *37*, 101391.

Haynos, A. F., Berg, K. C., Cao, L., Crosby, R. D., Lavender, J. M., Utzinger, L. M., Wonderlich, S. A., Engel, S. G., Mitchell, J. E., Le Grange, D., Peterson, C. B., & Crow, S. J. (2017). Trajectories of higher-and lower-order dimensions of negative and positive affect relative to restrictive eating in anorexia nervosa. *Journal of Abnormal Psychology*, *126*(5), 495.

Haynos, A. F., Koithan, E., & Hagan, K. E. (2022a). Learned industriousness as a translational mechanism in anorexia nervosa. *Nature Reviews Psychology*, 1–15.

Haynos, A. F., Lavender, J. M., Nelson, J., Crow, S. J., & Peterson, C. B. (2020). Moving towards specificity: A systematic review of cue features associated with reward and punishment in anorexia nervosa. *Clinical Psychology Review*, *79*, 101872.

Haynos, A. F., Widge, A. S., Anderson, L. M., & Redish, A. D. (2022b). Beyond description and deficits: How computational psychiatry can enhance an understanding of decision-making in anorexia nervosa. *Current Psychiatry Reports*, 1–11.

Howard, M., Gregertsen, E. C., Hindocha, C., & Serpell, L. (2020). Impulsivity and compulsivity in anorexia and bulimia nervosa: A systematic review. *Psychiatry Research*, *293*, 113354.

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, JAGS, and stan*.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470.

Matton, A., Goossens, L., Vervaet, M., & Braet, C. (2015). Temperamental differences between adolescents and young adults with or without an eating disorder. *Comprehensive Psychiatry*, *56*, 229–238.

Pedersen, M. L., & Frank, M. J. (2020). Simultaneous hierarchical bayesian parameter estimation for reinforcement learning and drift diffusion models: A tutorial and links to neural data. *Computational Brain & Behavior*, *3*, 458–471.

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review*, *24*, 1234–1251.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for raven's progressive matrices and vocabulary scales. Section 3: The standard progressive matrices*. Oxford Psychologists Press.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rotella, F., Mannucci, E., Gemignani, S., Lazzeretti, L., Fioravanti, G., & Ricca, V. (2018). Emotional eating and temperamental traits in eating disorders: A dimensional approach. *Psychiatry Research*, *264*, 1–8.

Schiff, S., Testa, G., Rusconi, M. L., Angeli, P., & Mapelli, D. (2021). Expectancy to eat modulates cognitive control and attention toward irrelevant food and non-food images in healthy starving individuals. A behavioral study. *Frontiers in Psychology*, *11*, 3902.

Shahar, N., Moran, R., Hauser, T. U., Kievit, R. A., McNamee, D., Moutoussis, M., … Dolan, R. J. (2019). Credit assignment to state-independent task representations and its relationship with model-based decision making. *Proceedings of the National Academy of Sciences*, *116*(32), 15871–15876.

Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: Insights from social neuroscience. *Nature Reviews Neuroscience*, *18*(9), 559–567.

Tchanturia, K., Davies, H., Reeder, C., & Wykes, T. (2010). *Cognitive remediation therapy for anorexia nervosa*. King's College London, University of London.

Trapp, W., Heid, A., Röder, S., Wimmer, F., & Hajak, G. (2022). Cognitive remediation in psychiatric disorders: State of the evidence, future perspectives, and some bold ideas. *Brain Sciences*, *12*(6), 683.

Wu, M., Brockmeyer, T., Hartmann, M., Skunde, M., Herzog, W., & Friederich, H.-C. (2014). Set-shifting ability across the spectrum of eating disorders and in overweight and obesity: A systematic review and meta-analysis. *Psychological Medicine*, *44*(16), 3365–3385.

Table 1.

Table of the Deviance Information Criterion (DIC) for the examined models.

Model	DIC
M1	90398.30
M2	89191.54
M3	89266.80
M4	88482.74
M5	85369.79
M6	84366.38
M7	82125.84

Figure 1.

Top. The probabilistic reversal-learning task is illustrated in a single trial. In each trial, subjects are presented with two figures and must determine the correct figure through trial-and-error feedback. Upon selection of a figure via left or right button press, feedback is provided in the form of a euro coin or a crossed euro coin. **Bottom:** The trial-by-trial proportion of choosing the image with the highest probability of reward in the first epoch computed for all participants.

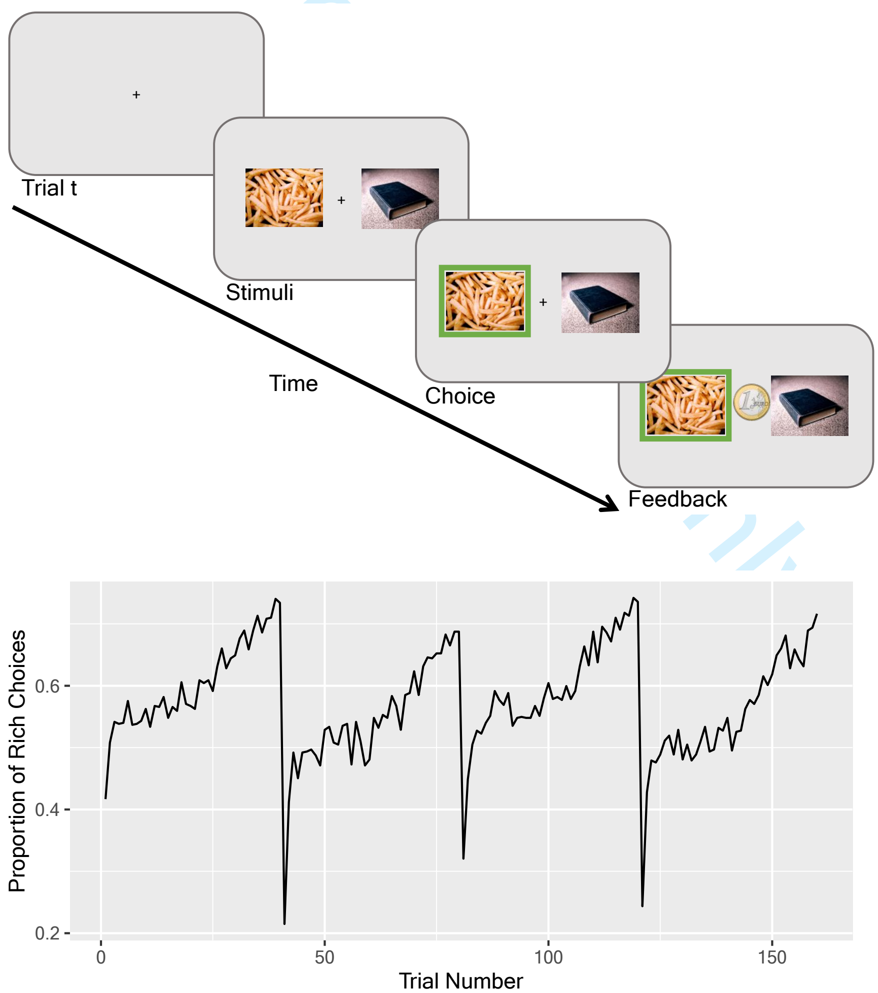


Figure 2.

Plots of the posterior distribution of the domain-specificity effect for parameter a ($a_{\text{food - related}}$ – $a_{\text{food - unrelated}}$) of the DDMRL across the four participants' groups.

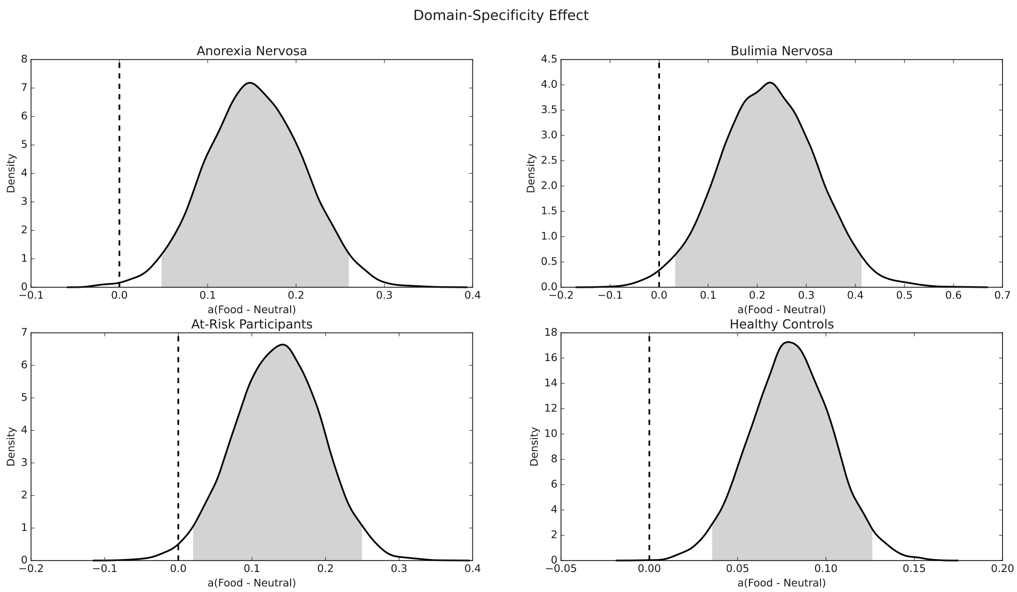


Figure 3.

Plots of the posterior distribution of the group effect for parameter α^+ $(\alpha^+_{group} - \alpha^+_{HC})$ of the DDMRL, for food-related (top row) and food-unrelated (bottom row) outcome-irrelevant information.

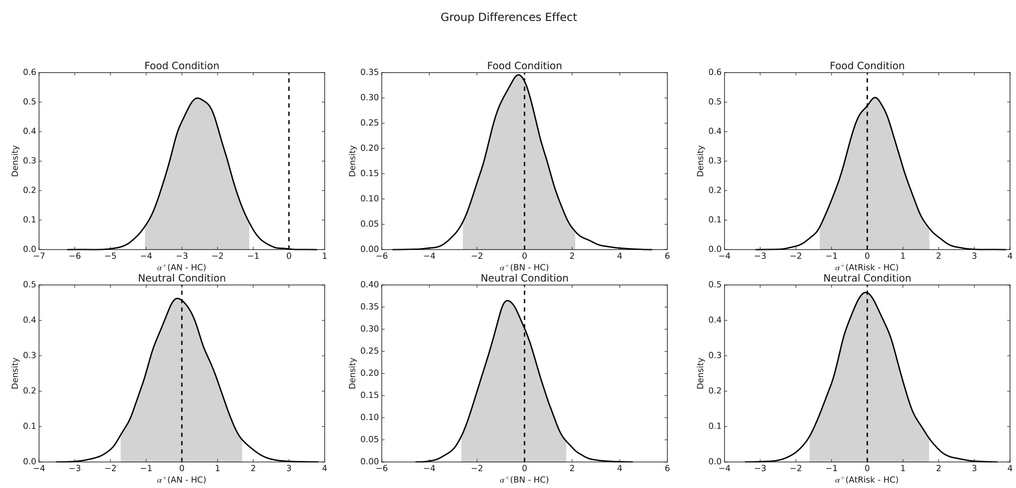


Figure 4.

Plots of the posterior distribution of the group effect for parameter α^- $(\alpha^-_{group} - \alpha^-_{HC})$ of the DDMRL, for food-related (top row) and food-unrelated (bottom row) outcome-irrelevant information.

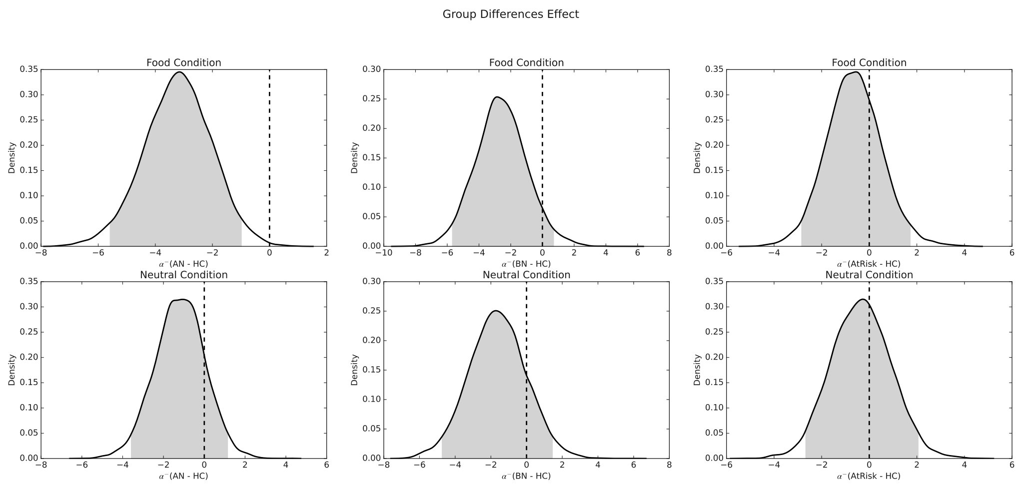


Figure 5.

Plots of the posterior distribution of the group effect ($a_{group} - a_{HC}$) for parameter a of the DDMRL, for food-related (top row) and food-unrelated (bottom row) outcome-irrelevant information.

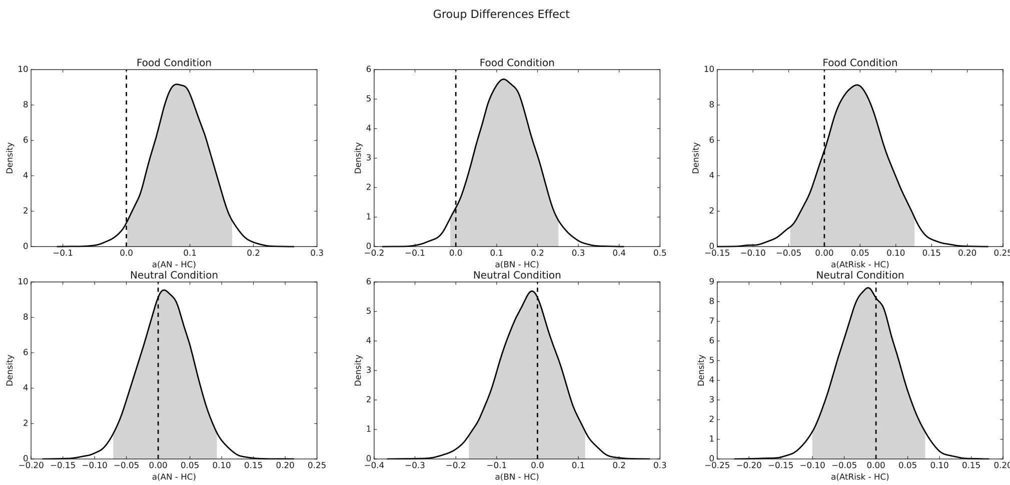
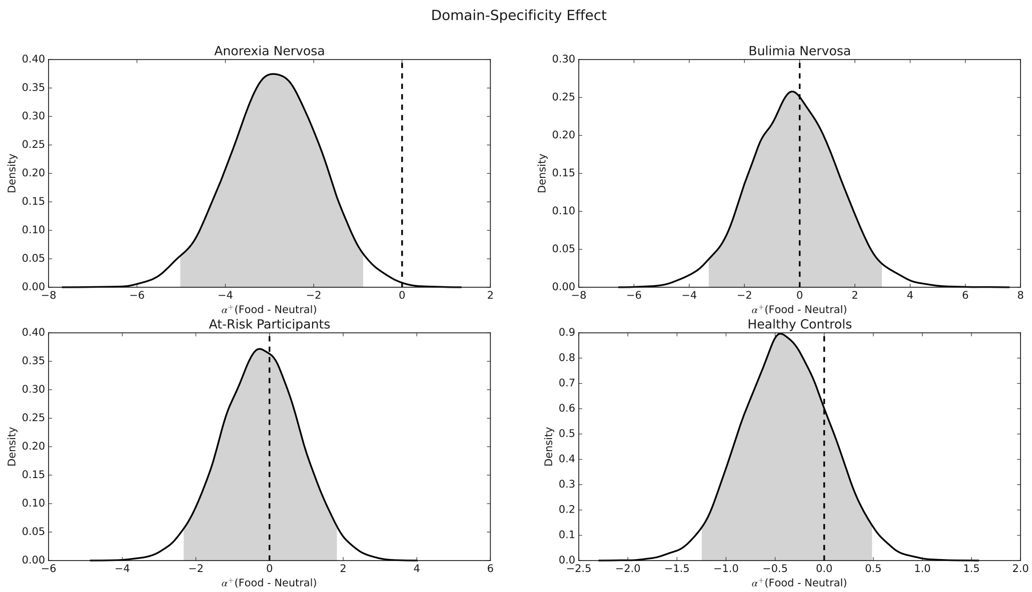


Figure 6.

Plots of the posterior distribution of the domain-specificity effect for parameter α^+ ($\alpha^+_{\text{food - related}} - \alpha^+_{\text{food - unrelated}}$) of the DDMRL across the four participants' groups.



Supplementary Information

When Food Becomes a Distraction: The Impact of Food-Related Information on Learning in Anorexia Nervosa

Contents

Participants sample size

To establish an appropriate sample size for our current study, we utilized data from a prior investigation that examined two distinct participant groups. Specifically, one group comprised 29 anorexic patients, while the other consisted of 124 healthy controls. These participants differed from those in our current study. In the prior study, each participant completed 160 trials per condition in a Probabilistic Reversal Learning (PRL) task, wherein the content of the image pairs presented in each trial was manipulated in a similar manner as in the current study. We determined that the experimental manipulation resulted in an effect size of 0.54 on the learning rate parameter, which is the primary focus of our current study.

To ascertain the required sample size necessary to detect a similar effect, we conducted a parameter recovery study using the methodology outlined by Pedersen & Frank (2020). Specifically, we simulated data from two groups of 30 participants with distinct α values (one lower and one higher) with a difference of 0.54. The remaining parameters of the reinforcement learning drift-diffusion model (namely, a, t, and v) were set to the values estimated from empirical data of the 29 anorexic patients in our prior study. For the simulation, we used the `hddm.generate.gen_rand_rlddm_data` function of the `hddm` module with following parameters:


```
subjects = 30
trials = 160

data = hddm.generate.gen_rand_rlddm_data(
    a=1.5,
    alpha=0.79, # or 0.25
    scaler=2.25,
    t=0.25,
    size=trials,
    subjs=subjects,
    p_upper=0.7,
    p_lower=0.3,
)
```

To estimate the parameters of the reinforcement learning drift-diffusion model (RLDDM) from the simulated data, we employed the `HDDMrl` function of the `hddm` module. To ensure the reliability of our estimates, we repeated this process 50 times. Notably, our results revealed that, in each iteration, the parameters for the lower and higher values of α were entirely separated. Consequently, our simulations demonstrate that the number of participants and trials included in our study were sufficient to detect the effect size on α observed in the previous study.

Although a parameter recovery study and a frequentist power analysis are fundamentally different techniques, we note that Bayesian methods place greater emphasis on parameter estimation rather than hypothesis testing. Therefore, it is reassuring to observe that, with the present number of participants and trials, the RLDDM model can successfully detect an effect size comparable to that previously identified in a distinct study featuring a distinct set of participants but a similar experimental manipulation.

Furthermore, we note that the sample sizes used in this study were comparable to those used in previous research (e.g., 32 subjects in Perez Santangelo et al., 2022; 41 women with AN in Foerde et al., 2021; 34 women with AN in Bernardoni et al., 2021).

Clinical Measurements

The *Eating Attitude Test-26* [EAT-26; Garner et al. (1982)] consists of 26 items assessing levels and types of eating disturbances in the past three months. The EAT-26 is characterized by three subscales: the Dieting Scale, the Bulimia and Food Preoccupation Scale and the Oral Control Scale. Scores ≥ 20 point out the presence of an eating disorder. Respondents are required to rate intensity associated with the items on a 6-point Likert scale (0 = never, rarely, sometimes; 3 = always). The Italian version of the EAT-26 demonstrated good psychometric properties (Dotti & Lazzari, 1998). In fact, Cronbach's alpha was high in an undergraduate sample for the Dieting scale (.87), for Bulimia and Food Preoccupation scale (.70), for Oral Control Scale (.62). Cronbach's alpha for the total scores was 0.86.

The *Body Shape Questionnaire-14* [BSQ-14; Dowson & Henderson (2001)] is a 14-item self-report scale assessing global body satisfaction over the past two weeks. Respondents are required to rate intensity of concerns about own appearance associated with the items on a 6-point Likert scale (1 = never, 6 = always). The Italian version of the BSQ-14 demonstrated good psychometric properties (Matera et al., 2013). In the present sample, $\omega = 0.978$.

The *Social Interaction Anxiety Scale* [SIAS; Mattick & Clarke (1998)] is a 20-item self-report questionnaire assessing social interaction anxiety. Respondents are required to rate intensity associated with the items on a 4-point Likert scale from 0 (not at all true) to 4 (extremely true). Higher scores denote greater social interaction anxiety levels. Both original version and the Italian version (Sica et al., 2007) show acceptable psychometric properties (in the present sample $\omega = 0.938$). Heimberg et al. (1992) have suggested a cut-off of 34 on the 20-item SIAS to denote a clinical level of social anxiety (32.3 for the Italian 19 item version).

The *Depression Anxiety Stress Scale-21* [DASS-21; Lovibond & Lovibond (1995)] is a 21-item self-report measure assessing depression, anxiety, and stress over the previous week. Items are rated on a 4-point scale ranging from 0 (did not apply to me at all) to 3 (applied to me very much). Both the original and the Italian version (Bottesi et al., 2015) demonstrate adequate reliability. In the present sample ω*“anxiety” = 0.875,* ω”depression” = 0.914, ω_“stress” = 0.899; for the total scale, $\omega = 0.945$.

The *Rosenberg Self-Esteem Scale* [RSES; Rosenberg (1965)] is a 10-item scale designed to assess person's overall self-esteem. It comprises five straightforwardly worded and five reverse-worded items each rated on a 4-point Likert scale ranging from 4 (strongly agree) to 1 (strongly disagree). Increased values indicate increased self-esteem. We used the Italian version of Prezza et al. (1997) which replicates the reliability of the original scale. In the present sample, $\omega = 0.949$.

The *Multidimensional Perfectionism Scale* [MPS-F; Frost et al. (1990)] is a 35-item assessing perfectionism tendencies. According to Stober (1998), MPS-F is composed of four underlying factors: Concerns over Mistakes and Doubts (CMD), Parental Expectations and Criticism (PEC), Personal Standards (PS), and Organization (O). Both the original MPS-F and the

Italian version (Lombardo, 2008) demonstrated adequate reliability. In the present sample, ω*"CMD" = 0.919,* ω"PS" = 0.851, ω*"PEPC" = 0.946,* ω"OR" = 0.931; for the total scale, $\omega = 0.932$.

In this study, all 60 items of the *Raven's Standard Progressive Matrices* (SPM) were utilized, which is a well-established measure of cognitive ability (J. Raven et al., 2000). The SPM has been found to possess robust psychometric properties, as evidenced by its high test-retest reliability coefficients ranging from 0.83 to 0.93, and construct validity coefficients ranging from 0.81 to 0.94 (J. Raven et al., 2000).

Demographic and self-report measures

Age

In our initial statistical analysis, we investigated if there were any differences in age among the groups (AN = Anorexia Nervosa, BN = Bulimia Nervosa, HC = Healthy Controls, RI = participants who were at risk of developing eating disorders but had not received a formal diagnosis).

The mean age and standard devitation by group is shown below.

```
## group_by: one grouping variable (group)
```

```
## summarize: now 4 rows and 3 columns, ungrouped
```

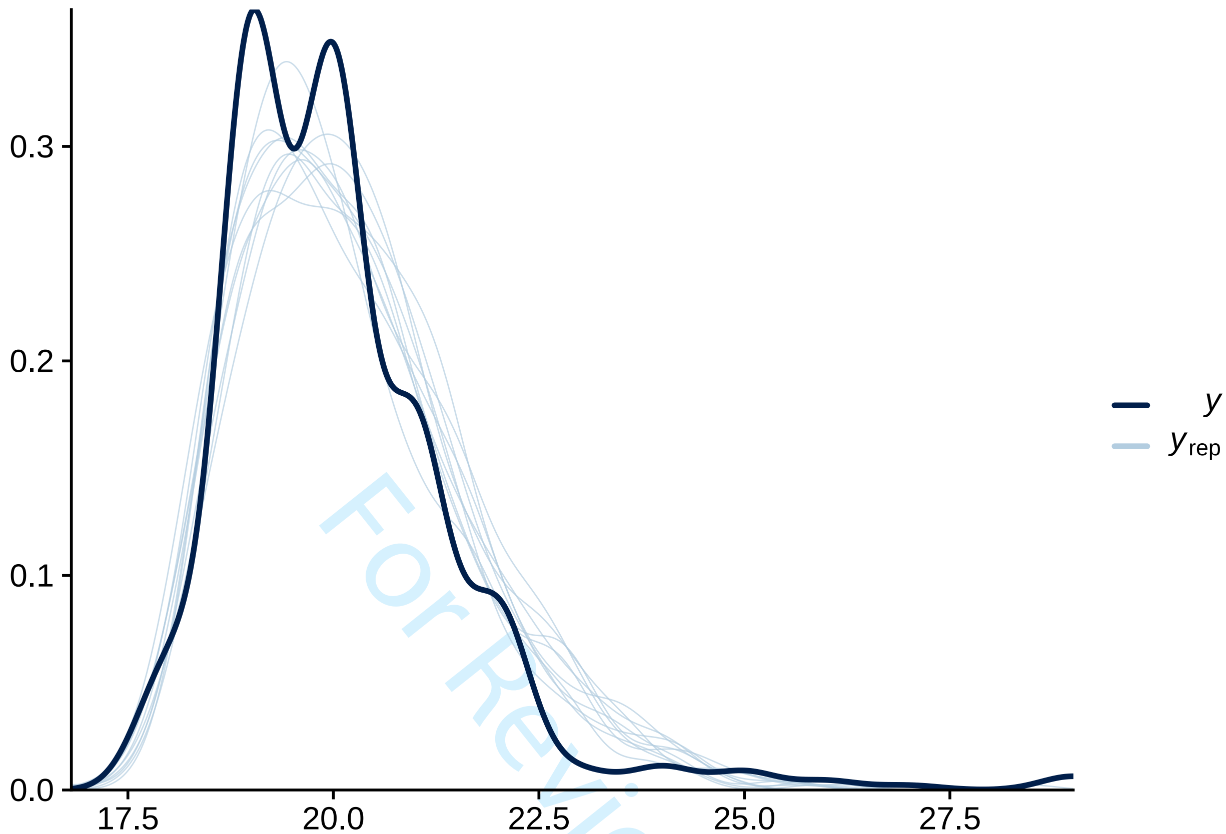
group	Age	SD
HC	19.77	1.06
AN	21.18	2.41
BN	20.39	1.88
RI	20.36	1.44

We used a Bayesian regression model to examine the age differences among groups.

```
m1 <- brm(
  age ~ group,
  data = quest_param_df,
  family = skew_normal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

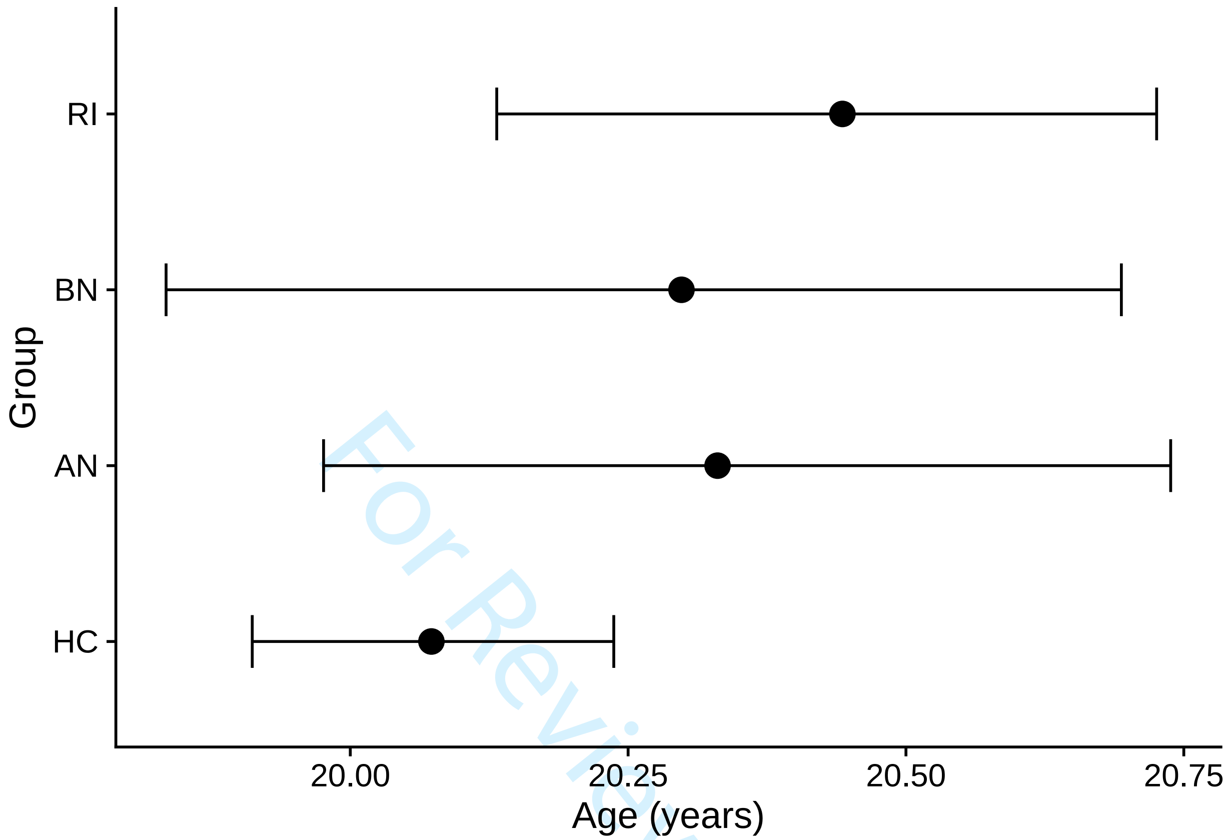


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	20.074	0.083	19.911	20.236
groupAN	0.265	0.234	-0.163	0.738
groupBN	0.217	0.223	-0.218	0.649
groupRI	0.364	0.170	0.013	0.677

Predicted effect of group on age



Interpretation

The 95% credibility intervals for the difference in age between each group and the HC baseline included zero, indicating that there were no credible differences in age among the groups.

Body Mass Index (BMI)

The average Body Mass Index (BMI) values and standard deviations for each group are reported below.

```
## group_by: one grouping variable (group)
```

```
## summarize: now 4 rows and 3 columns, ungrouped
```

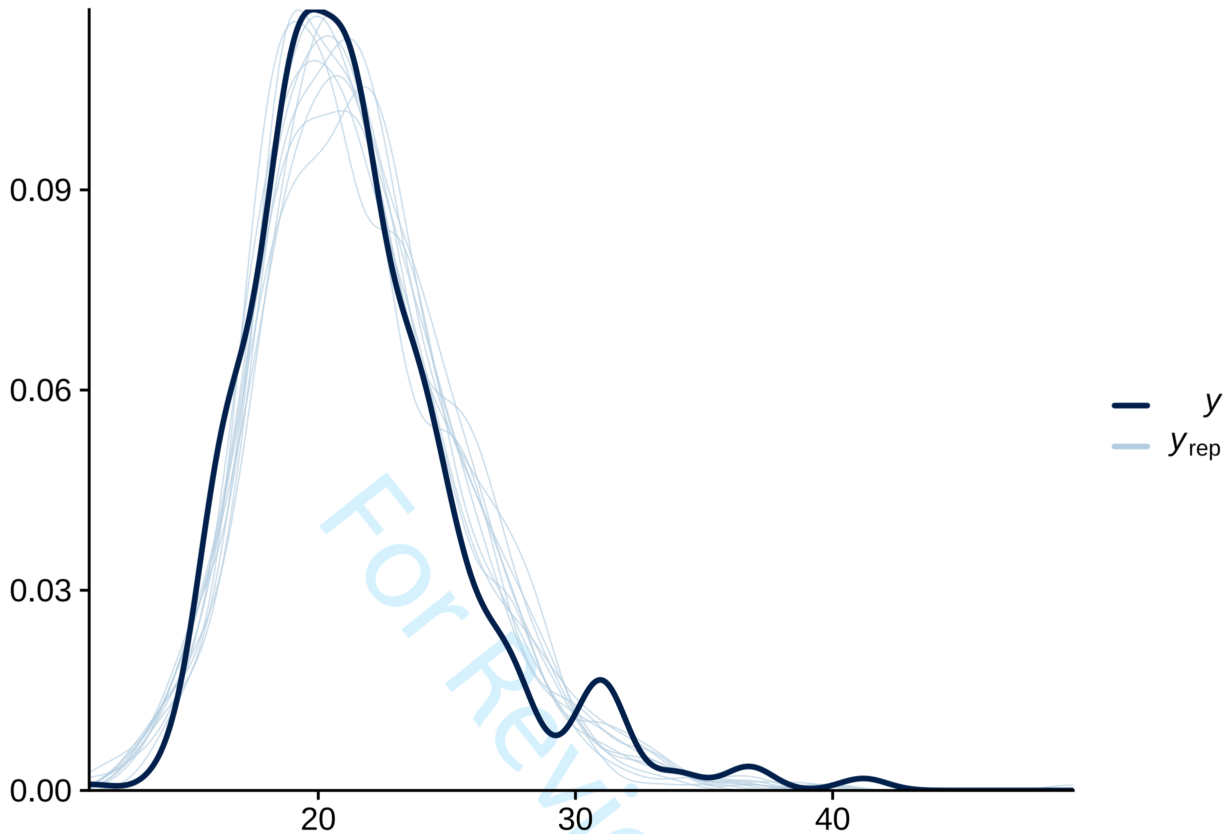
group	BMI	SD
HC	21.62	3.03
AN	16.88	1.55
BN	30.09	5.47
RI	22.41	4.79

We used a Bayesian regression model to examine the BMI differences among groups.

```
m2 <- brm(
  bmi ~ group,
  data = quest_param_df,
  family = skew_normal(),
  iter = 4000,
  cores = 4,
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

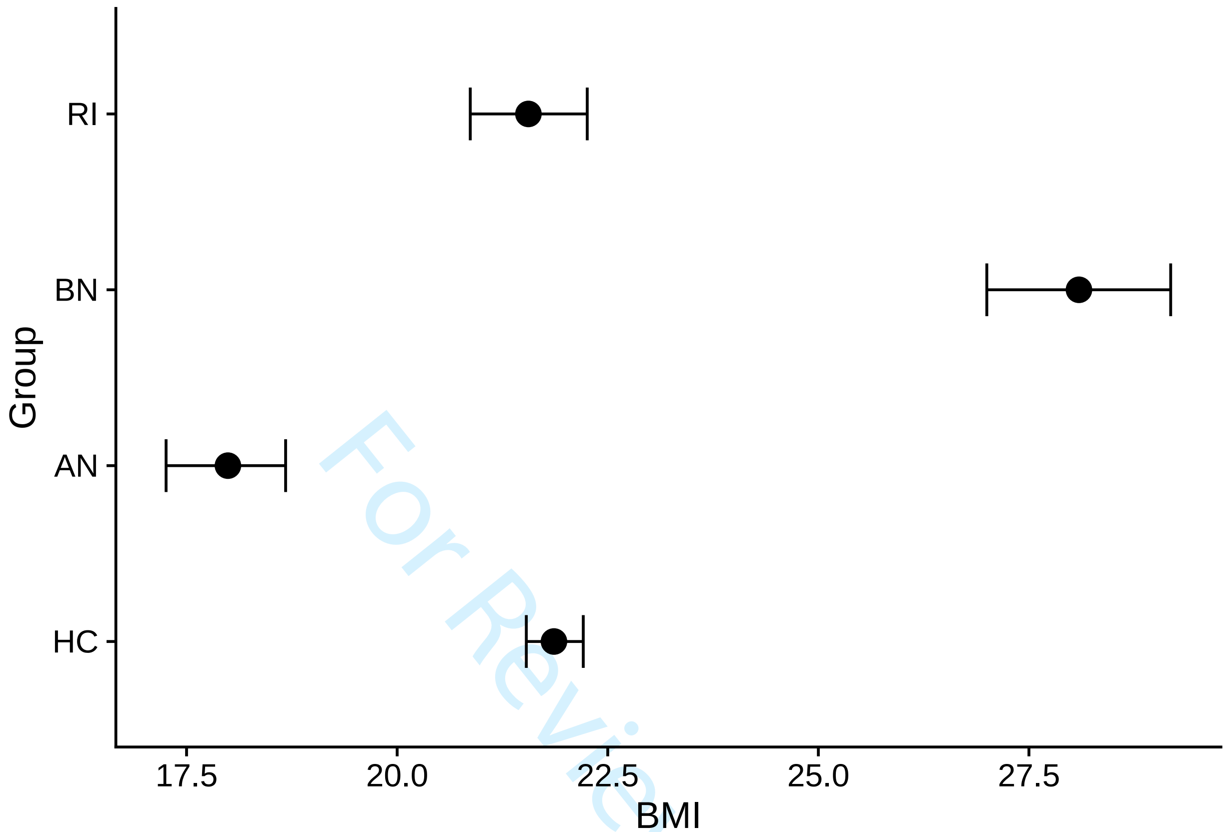


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	21.864	0.173	21.517	22.191
groupAN	-3.877	0.376	-4.642	-3.158
groupBN	6.232	0.580	5.082	7.350
groupRI	-0.304	0.386	-1.076	0.444

Predicted effect of group on BMI



Interpretation

The 95% credibility intervals for the difference in age between each group and the HC baseline do not include zero, indicating credible BMI differences between the AN and the HC groups, and between the BN and the HC groups. The at-risk and HC groups did not differ in terms of their average BMI values.

Rosenberg Self-Esteem scale

The average score of the Rosenberg scale as a function of group is shown below.

```
## distinct: removed 219 rows (46%), 259 rows remaining
```

```
## group_by: one grouping variable (group)
```

```
## summarize: now 4 rows and 3 columns, ungrouped
```

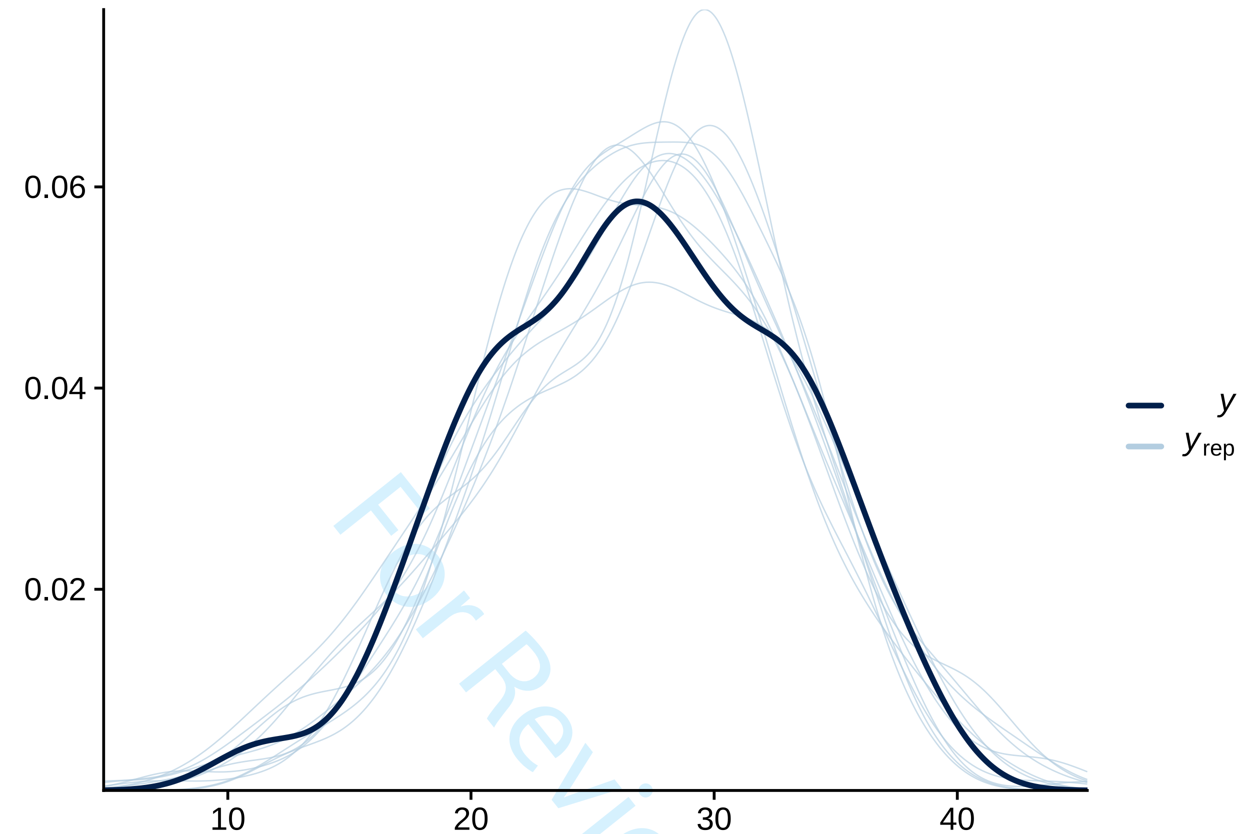
group	Rosenberg	SD
HC	28.80	5.50
AN	22.42	5.41
BN	20.91	4.91
RI	22.80	5.46

We used a Bayesian regression model to examine the differences in Rosenberg self-esteem scores among groups.

```
m3 <- brm(
  ros_tot ~ group,
  data = ros_df,
  family = skew_normal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

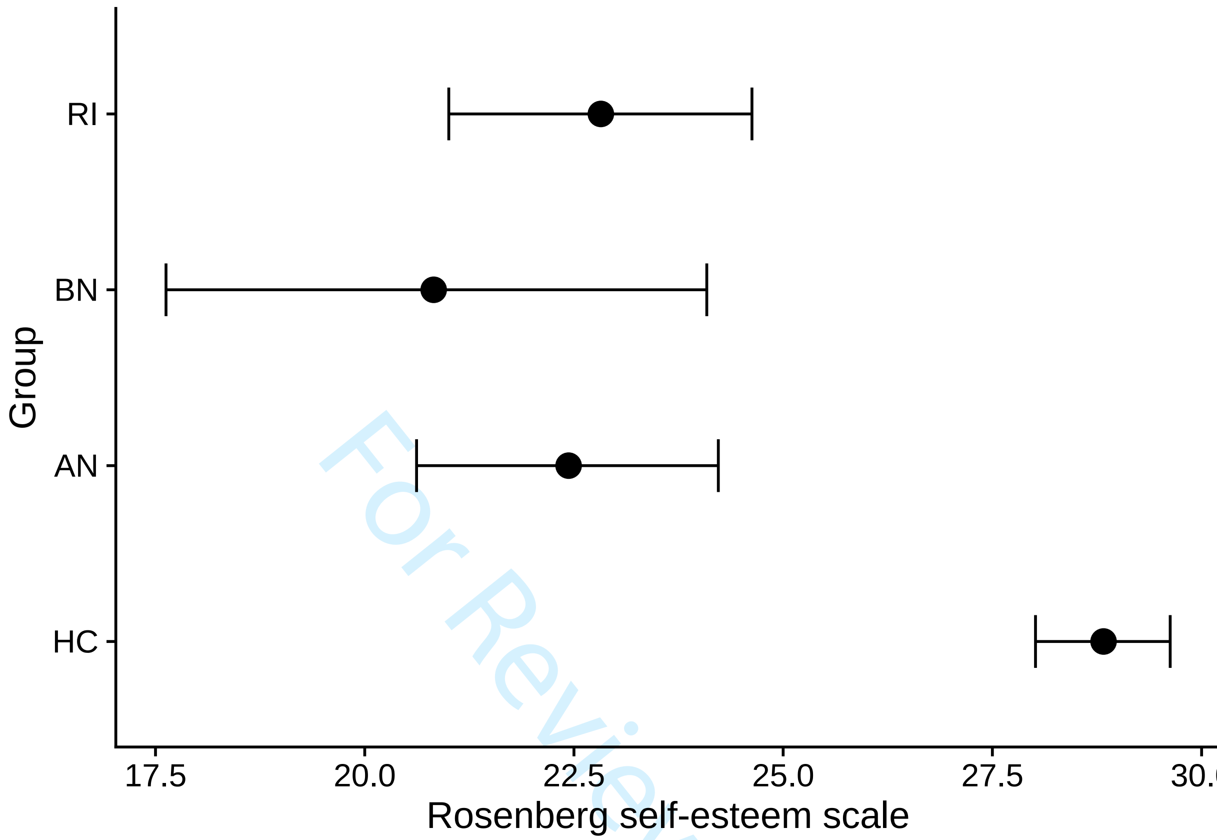


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	28.823	0.409	28.015	29.627
groupAN	-6.389	0.990	-8.371	-4.429
groupBN	-8.010	1.684	-11.390	-4.778
groupRI	-6.011	1.009	-7.981	-4.081

Predicted effect of group on the Rosenberg self-esteem scores



Interpretation

The 95% credibility intervals for the difference in the Rosenberg self-esteem scores between each group and the HC baseline do not include zero, indicating credible differences in the Rosenberg self-esteem scores between the HC and the other groups. The inferior self-esteem of those afflicted with AN compared to the HCs, along with the analogous level of self-esteem between individuals with AN and those with BN, replicate the results of a recent comprehensive overview of the current findings concerning self-esteem in patients with AN (Kaestner et al., 2019).

Body Shape Questionnaire-14 (BSQ-14)

The average Body Shape Questionnaire-14 score as a function of group is shown below.

```
## distinct: removed 219 rows (46%), 259 rows remaining
## group_by: one grouping variable (group)
## summarize: now 4 rows and 3 columns, ungrouped
```

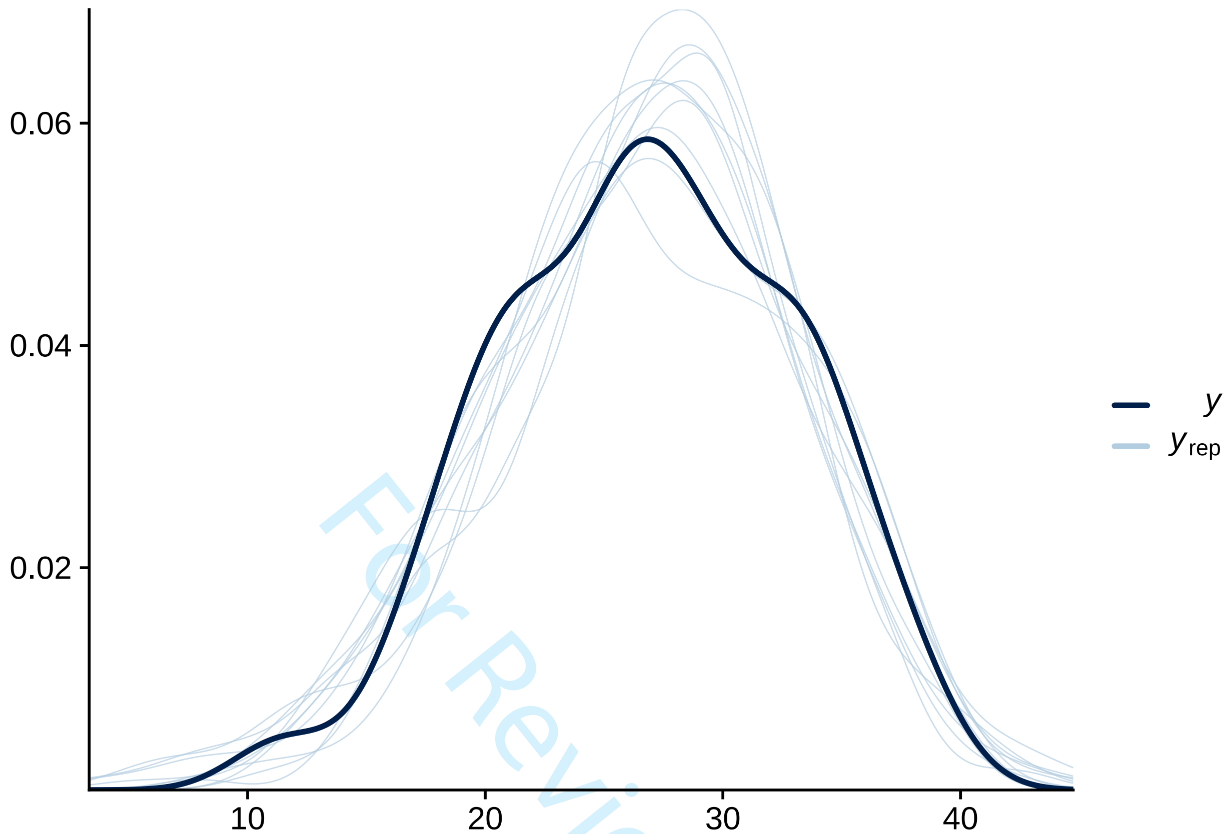
group	BSQ_14	SD
HC	89.90	32.96
AN	145.44	32.58
BN	154.77	33.60
RI	152.79	35.07

We used a Bayesian regression model to examine the BSQ-14 differences among groups.

```
m4 <- brm(
  ros_tot ~ group,
  data = ros_df,
  family = skew_normal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

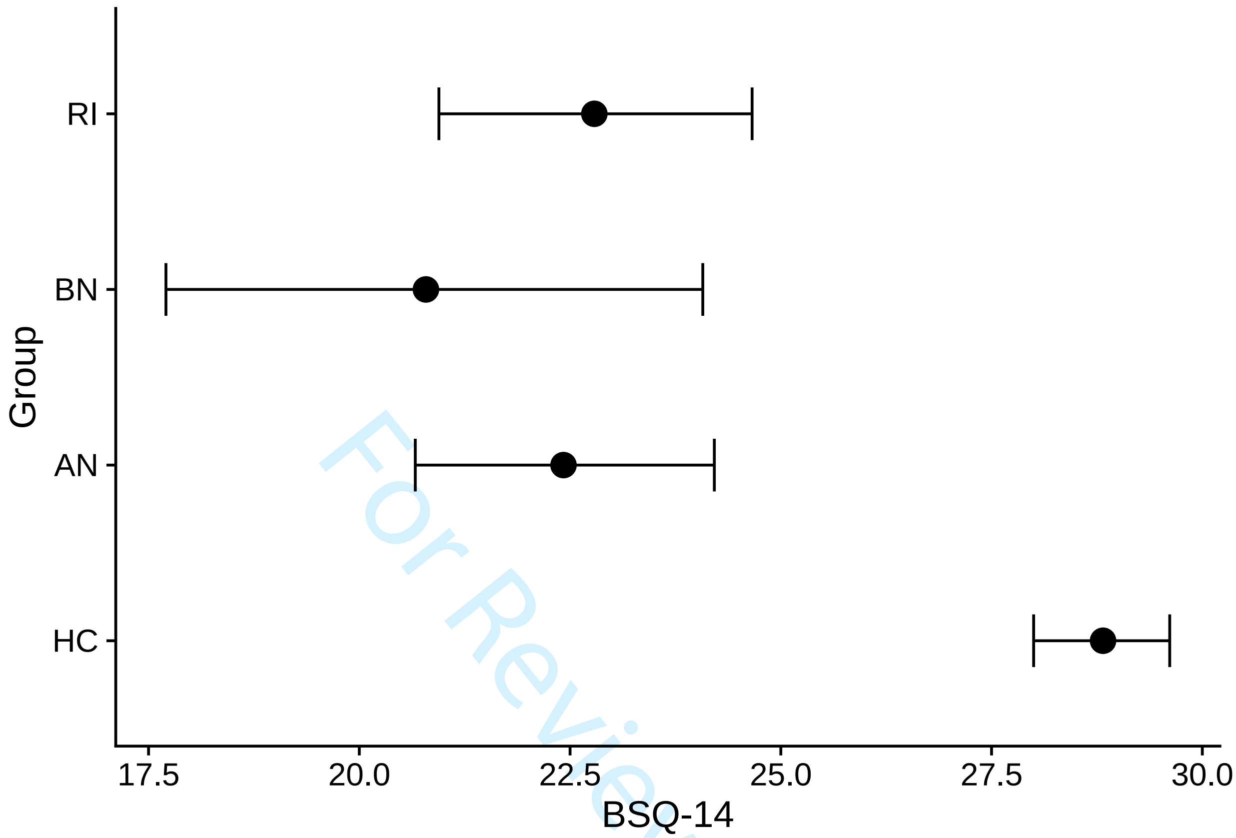


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	28.816	0.415	28.006	29.617
groupAN	-6.388	0.992	-8.304	-4.428
groupBN	-8.003	1.679	-11.241	-4.714
groupRI	-6.012	1.019	-8.088	-4.043

Predicted effect of group on the BSQ-14 scores



Interpretation

Our analysis revealed that the 95% credibility intervals for the difference in BSQ-14 scores between each group and the healthy control (HC) baseline do not encompass zero, indicating credible differences in the BSQ-14 scores between the HC and the other groups. Therefore, this results replicate the well-known finding that body image disturbances are a hallmark of both AN and BN disorders (Cooper et al., 1987).

Depression Anxiety Stress Scale-21 (DASS-21)

```
## distinct: removed 219 rows (46%), 259 rows remaining
```

Average values of the scores on the Stress, Anxiety, and Depression of the DASS-21 scale are shown below.

```
## group_by: one grouping variable (group)
```

```
## summarize: now 4 rows and 4 columns, ungrouped
```

group	Stress	Anxiety	Depression
HC	9.17	5.30	6.78
AN	13.25	8.56	10.86
BN	10.73	6.73	11.91
RI	12.66	7.69	11.37

Standard deviations of the scores on the Stress, Anxiety, and Depression sub-scales are shown below.

```
## group_by: one grouping variable (group)
```

```
## summarize: now 4 rows and 4 columns, ungrouped
```

group	SD_stress	SD_anxiety	SD_depression
HC	9.17	5.30	6.78
AN	13.25	8.56	10.86
BN	10.73	6.73	11.91
RI	12.66	7.69	11.37

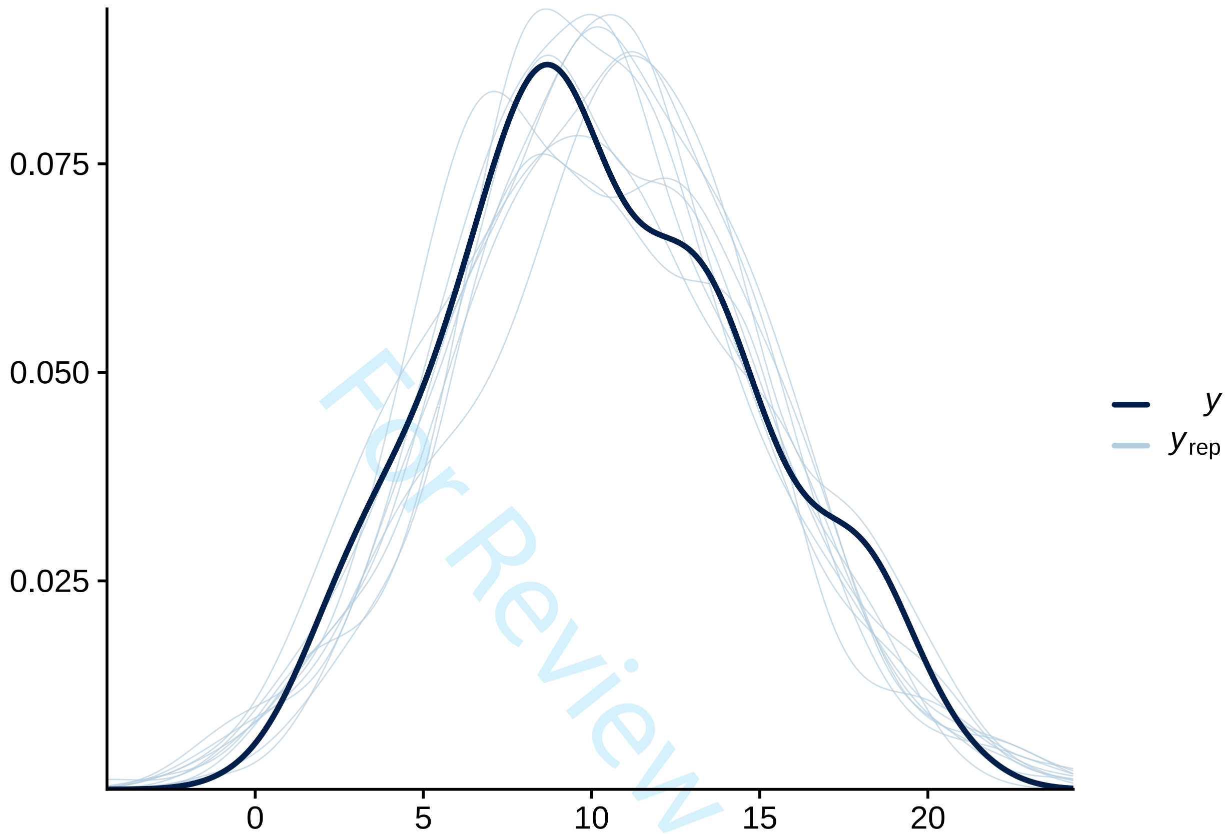
Stress

We used a Bayesian regression model to examine the DASS-21 Stress differences among groups.

```
m5_s <- brm(
  dass21_s ~ group,
  data = dass_df,
  family = gaussian(),
  iter = 4000,
  cores = 4,
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

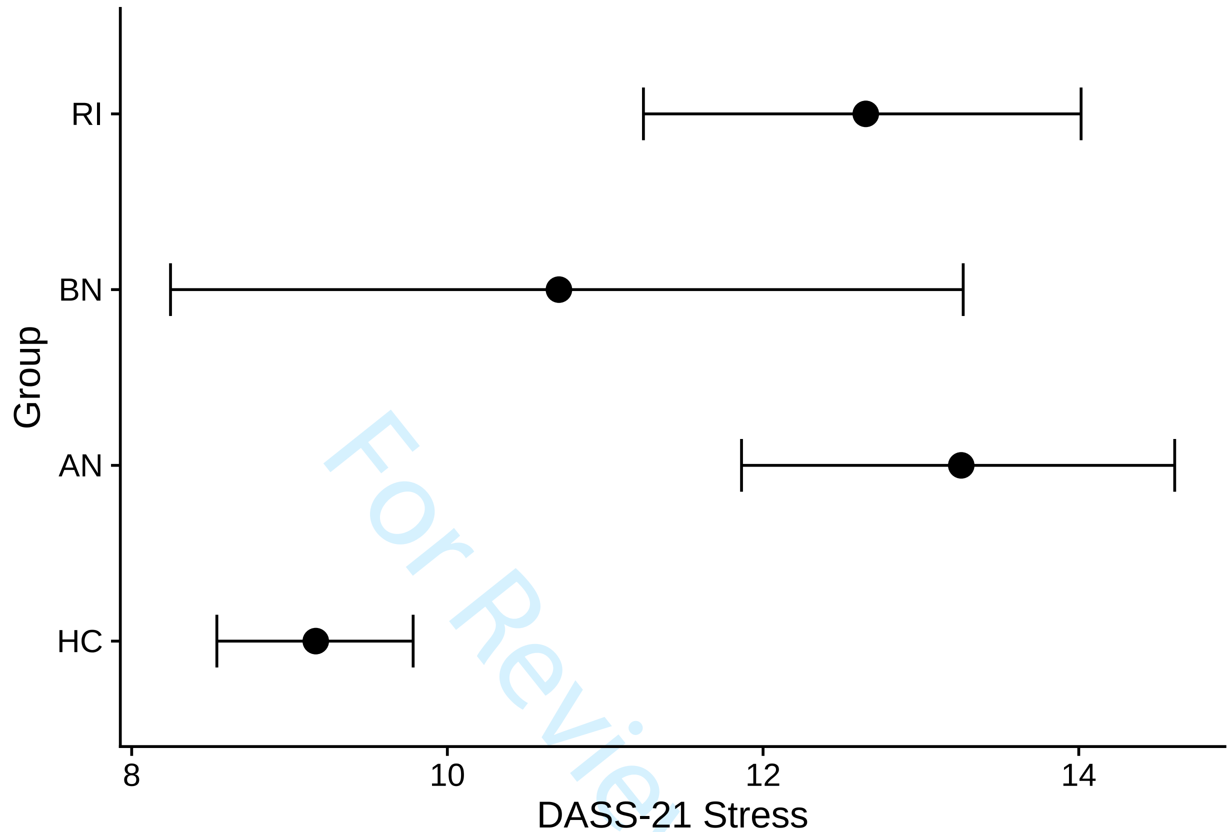


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	9.167	0.318	8.530	9.765
groupAN	4.079	0.776	2.583	5.574
groupBN	1.560	1.320	-0.867	4.269
groupRI	3.480	0.782	1.931	4.976

Predicted effect of group on the DASS-21 Stress scores



Interpretation

The 95% credibility intervals for the difference in the DASS-21 Stress scores between the AN and RI groups, on the one side, and the HC baseline, on the other, do not include zero, indicating credible differences in the DASS-21 Stress between the HC and the other two groups. We found no credible difference in the average DASS-21 Stress score between the HC and BC groups. The results replicate the susceptibility of individuals with AN to stress when compared to the general population (Guarda et al., 2015).

Anxiety

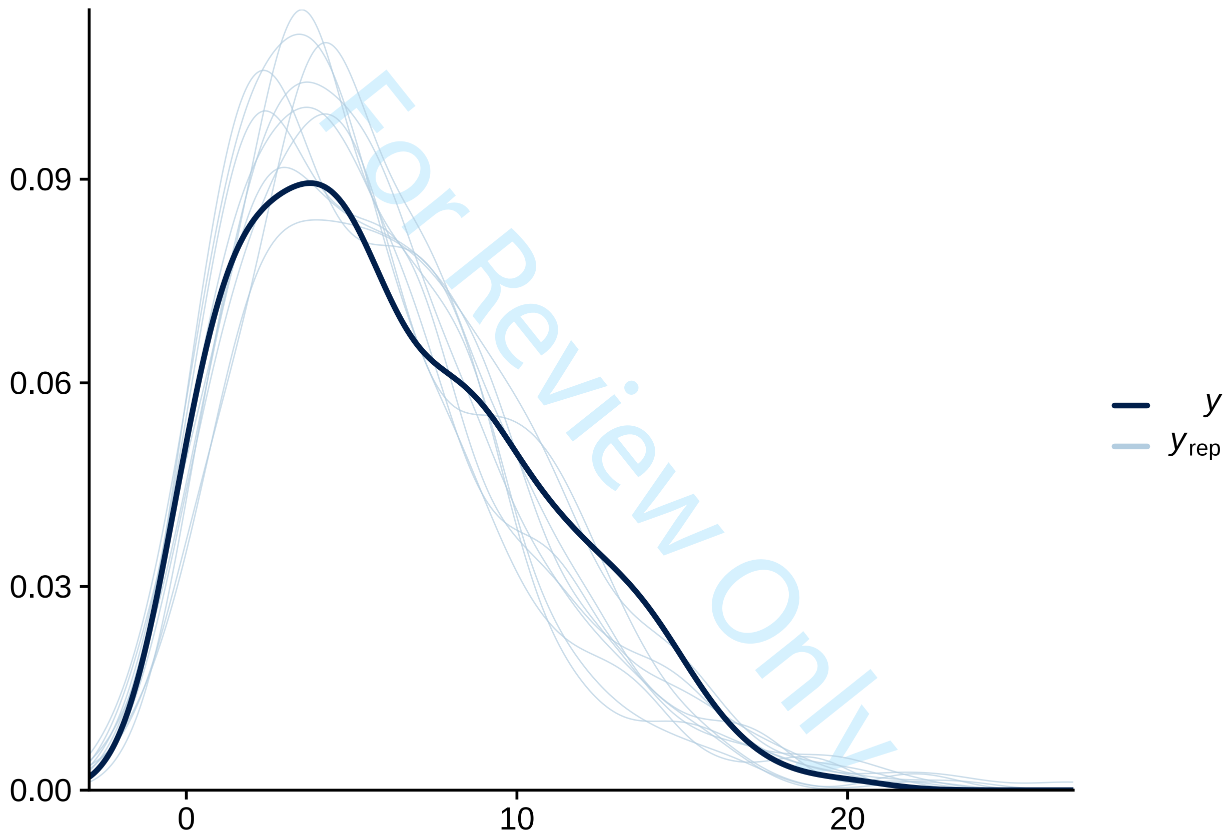
We used a Bayesian regression model to examine the DASS-21 Anxiety differences among groups.

```
m5_a <- brm(
  dass21_a ~ group,
  data = dass_df,
  family = skew_normal(),
  iter = 4000,
```

```
  cores = 4,
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

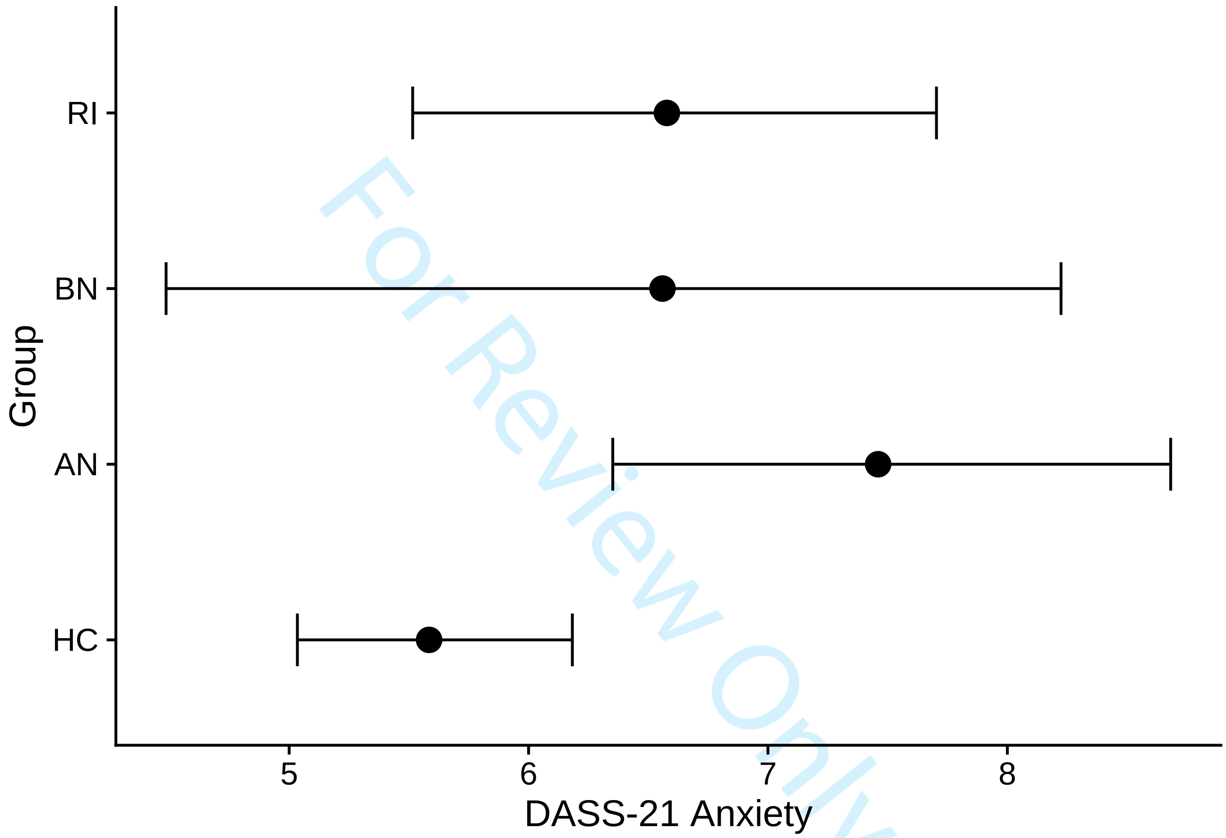


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	5.589	0.294	5.013	6.160
groupAN	1.885	0.632	0.680	3.138
groupBN	0.915	0.966	-1.014	2.811
groupRI	0.995	0.579	-0.110	2.155

Predicted effect of group on the DASS-21 Anxiety scores



Interpretation

The 95% credibility intervals for the difference in the DASS-21 Anxiety scores between the AN and the HC baseline do not include zero, indicating credible differences in the DASS-21 Anxiety between the two groups. We found no credible difference in the average DASS-21 Anxiety score between the HC and BC groups, nor between the HC and the RI groups. The results confirm an increase in anxiety levels among individuals with AN when contrasted with the general population (Swinbourne & Touyz, 2007).

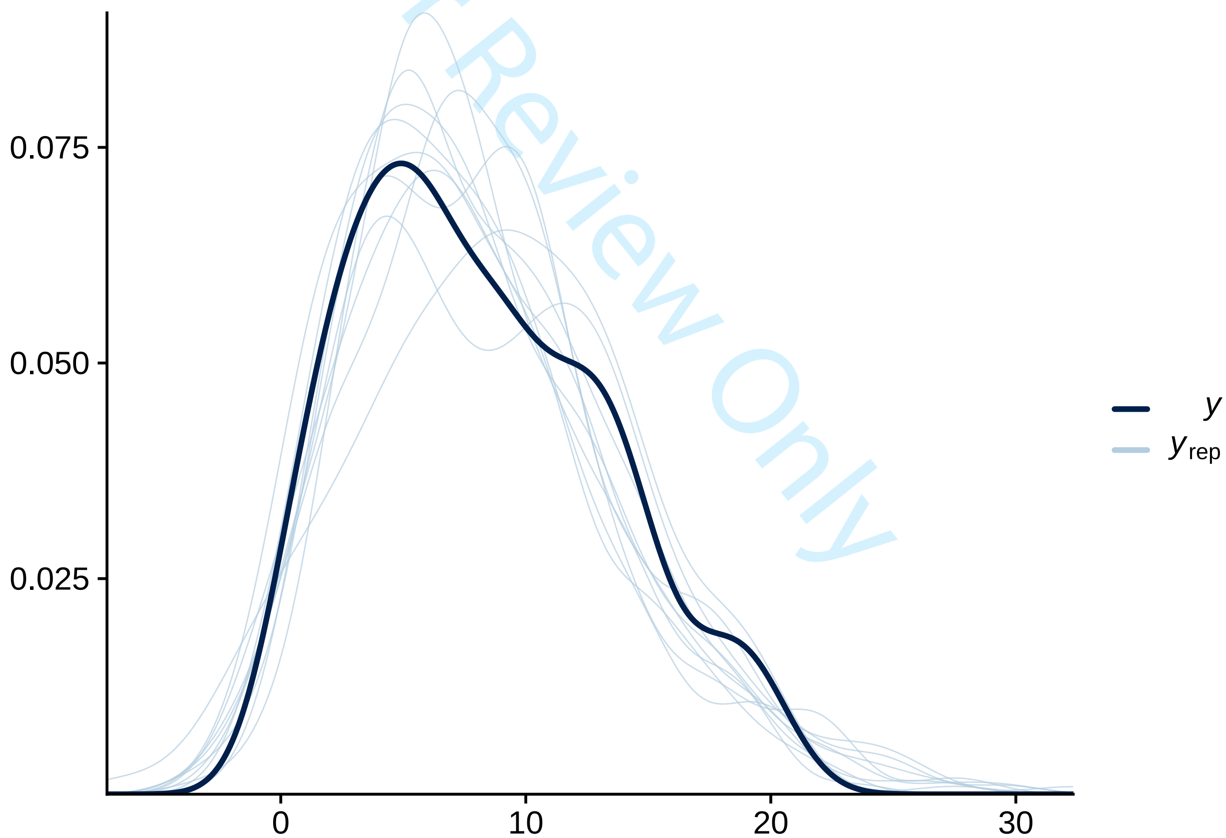
Depression

We used a Bayesian regression model to examine the DASS-21 Depression differences among groups.

```
m5_d <- brm(
  dass21_d ~ group,
  data = dass_df,
  family = skew_normal(),
  iter = 4000,
  cores = 4,
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

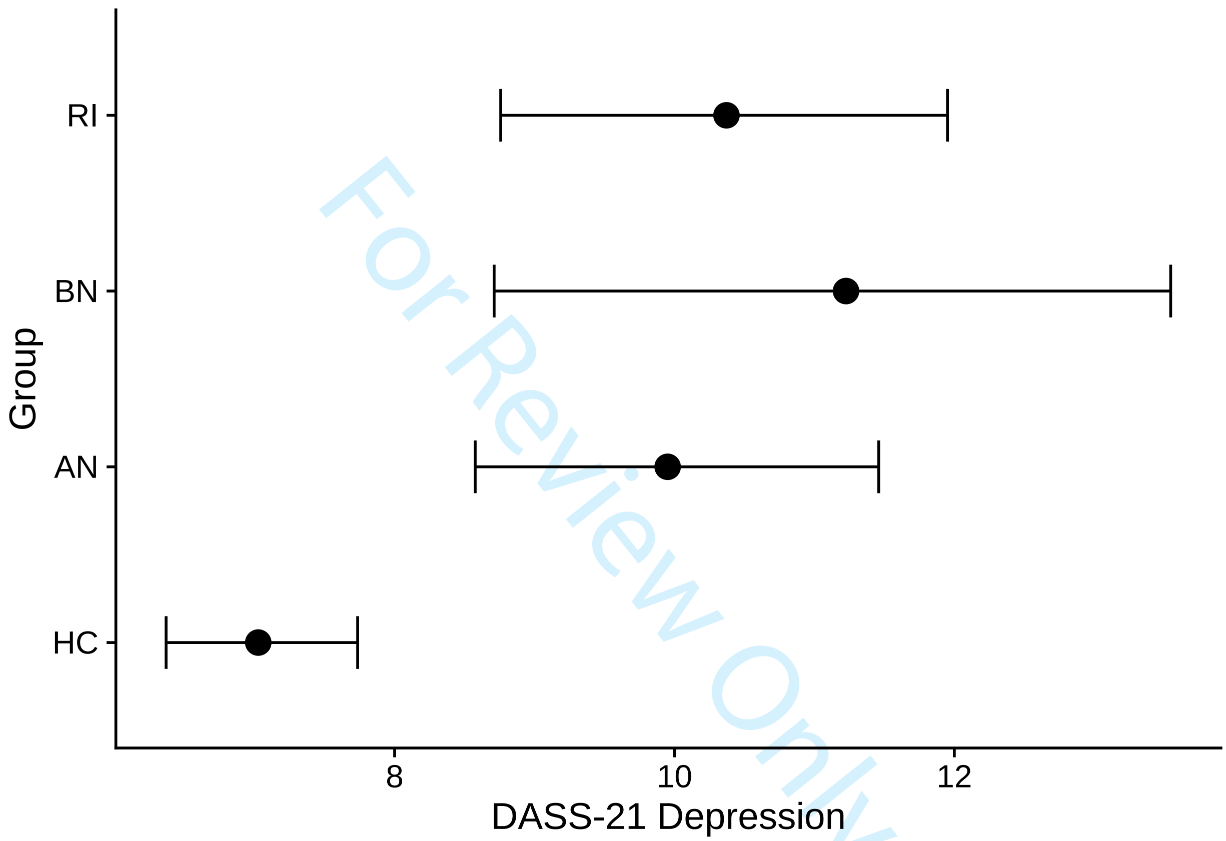


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	7.033	0.353	6.381	7.748
groupAN	2.942	0.804	1.419	4.534
groupBN	4.154	1.248	1.565	6.578
groupRI	3.331	0.880	1.590	5.038

Predicted effect of group on the DASS-21 Depression scores



Interpretation

The 95% credibility intervals for the difference in the DASS-21 Depression scores between the AN, BN, and RI groups, on the one side, and the HC baseline, on the other, do not include zero, indicating credible differences in the DASS-21 Depression between the baseline group and the other groups. Our results are in agreement with previous research indicating a positive association between symptoms of AN and depressive symptoms, even during treatment of AN patients (Pleple et al., 2021).

Social Interaction Anxiety Scale (SIAS)

```
## distinct: removed 219 rows (46%), 259 rows remaining
```

```
## group_by: one grouping variable (group)
```

```
## summarize: now 4 rows and 3 columns, ungrouped
```

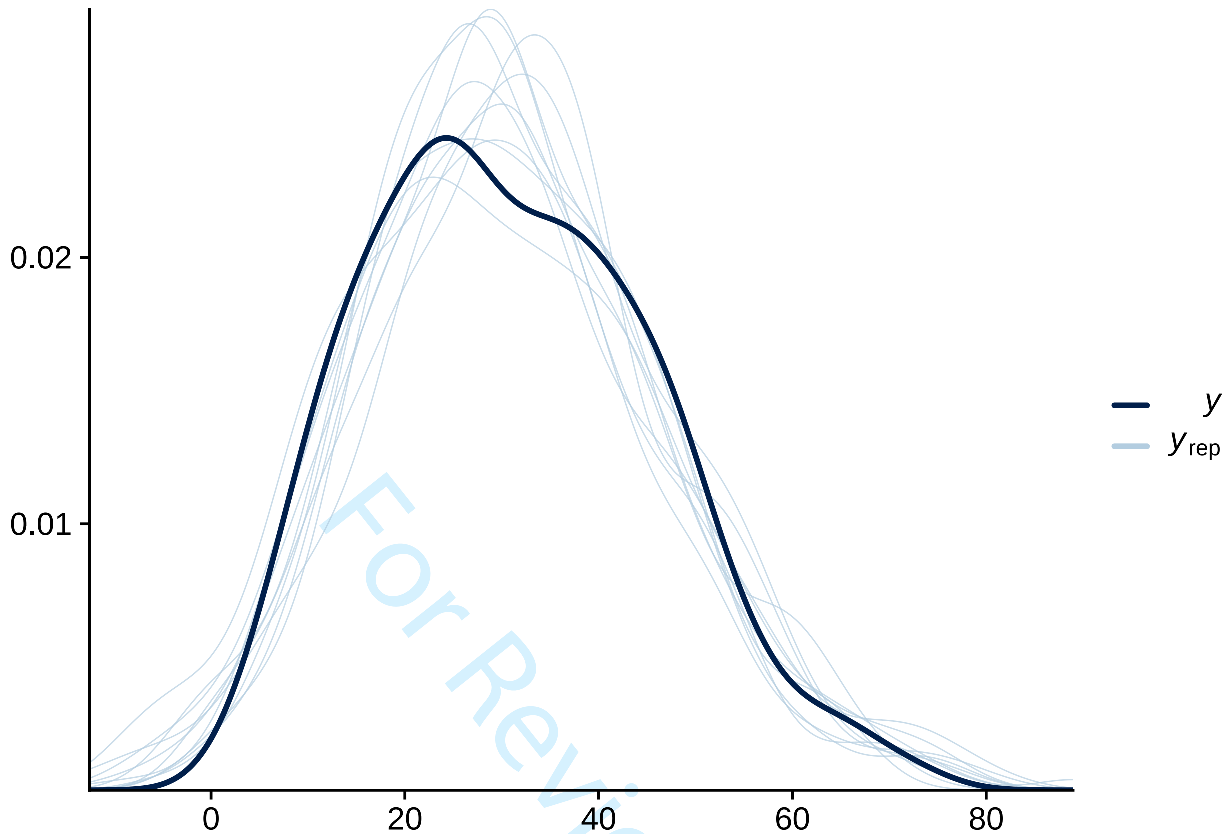
group	SIAS	SD
HC	28.03	13.07
AN	37.31	15.44
BN	34.64	15.19
RI	37.31	17.09

We used a Bayesian regression model to examine the SIAS score differences among groups.

```
m6 <- brm(
  sias_tot ~ group,
  data = sias_df,
  family = skew_normal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

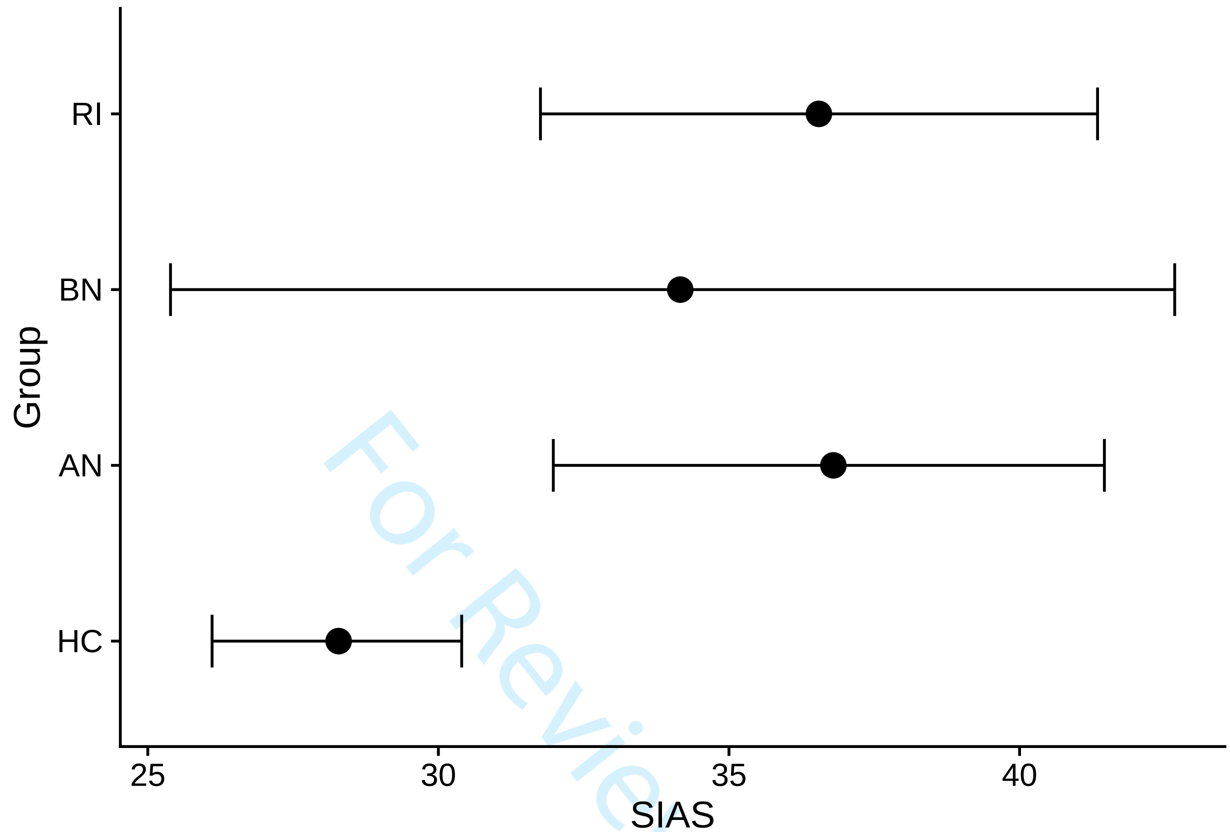


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	28.275	1.097	25.966	30.258
groupAN	8.509	2.682	3.060	13.524
groupBN	5.852	4.562	-2.889	14.750
groupRI	8.270	2.741	2.929	13.604

Predicted effect of group on the SIAS scores



Interpretation

The 95% credibility intervals for the difference in the SIAS scores between the AN and RI groups, on the one side, and the HC baseline, on the other do not include zero, indicating credible differences in the SIAS scores between the HC and these two groups. We found no credible difference in the average SIAS scores between the BN and the HC groups. Our findings provide support for the notion that social anxiety represents a significant contributory factor in AN (Grabhorn et al., 2006).

Multidimensional Perfectionism Scale (MPS)

The average scores on the sub-scales of Concerns over Mistakes and Doubts (CMD), Parental Expectations and Criticism (PEC), Personal Standards (PS), and Organization (O) of the MPS as a function of group are shown below.

```
## distinct: removed 219 rows (46%), 259 rows remaining

## group_by: one grouping variable (group)

## summarize: now 4 rows and 5 columns, ungrouped
```

group	MPS_cmd	MPS_ps	MPS_pepc	MPS_or
HC	38.28	22.01	21.20	22.23
AN	45.19	25.58	21.72	23.47
BN	44.91	23.55	21.36	23.00
RI	48.74	25.89	25.69	22.80

The standard deviations on the sub-scales of Concerns over Mistakes and Doubts (CMD), Parental Expectations and Criticism (PEC), Personal Standards (PS), and Organization (O) of the MPS as a function of group are shown below.

```
## group_by: one grouping variable (group)

## summarize: now 4 rows and 5 columns, ungrouped
```

group	SD_cmd	SD_ps	SD_pepc	SD_or
HC	8.66	4.97	7.09	4.73
AN	8.43	5.52	8.00	5.63
BN	9.43	5.39	5.35	5.08
RI	8.66	5.89	7.61	5.76

Concerns over Mistakes and Doubts

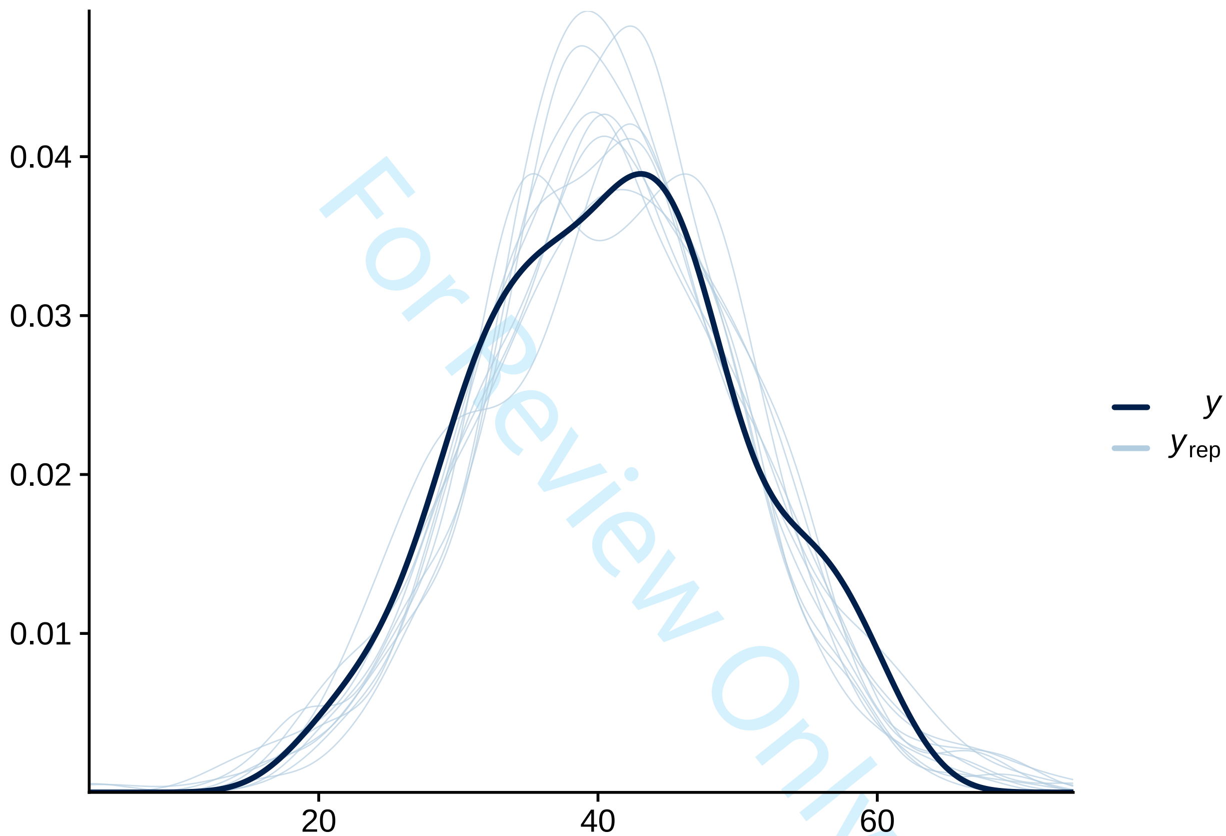
We used a Bayesian regression model to examine the MPS CMD score differences among groups.

```
m7_cmd <- brm(
  mps_cmd ~ group,
  data = mps_df,
  family = gaussian(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
```

```
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

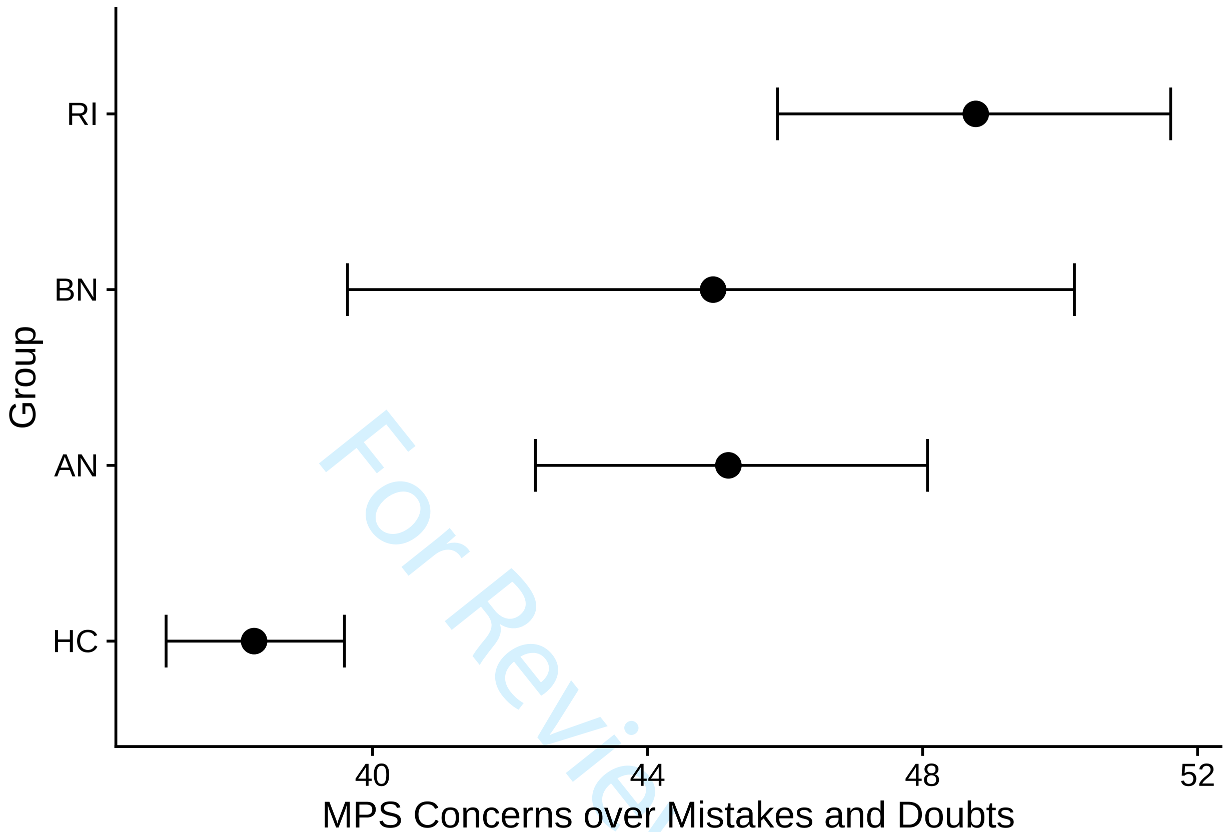


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	38.276	0.656	37.052	39.636
groupAN	6.914	1.592	3.731	9.952
groupBN	6.652	2.768	1.170	11.945
groupRI	10.491	1.608	7.399	13.623

Predicted effect of group on the MPS Concerns over Mistakes and Doubts



Interpretation

The 95% credibility intervals for the difference in the MPS-CMD scores between the HC and the other groups do not include zero, indicating credible differences in the MPS-CMD scores between the HC and the other groups.

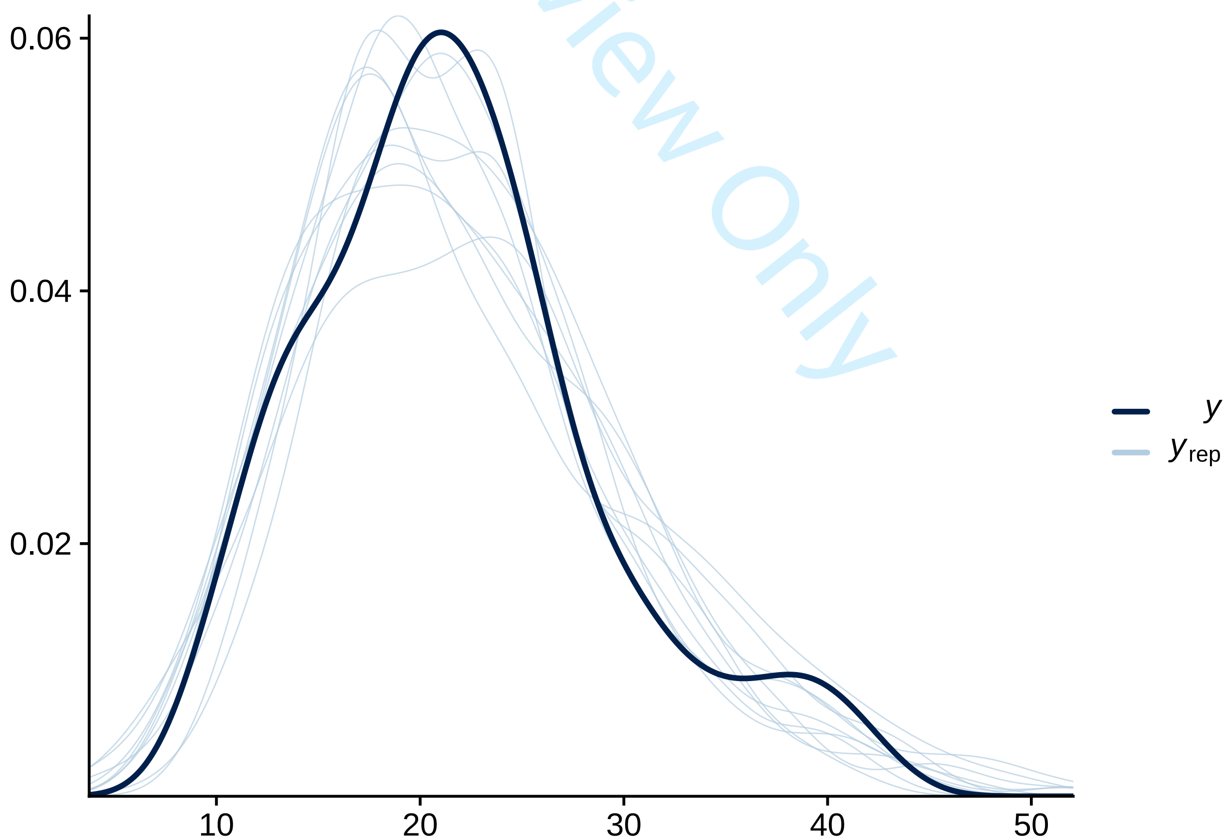
Parental Expectations and Criticism

We used a Bayesian regression model to examine the MPS PEC score differences among groups.

```
m7_pec <- brm(
  mps_pepc ~ group,
  data = mps_df,
  family = skew_normal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

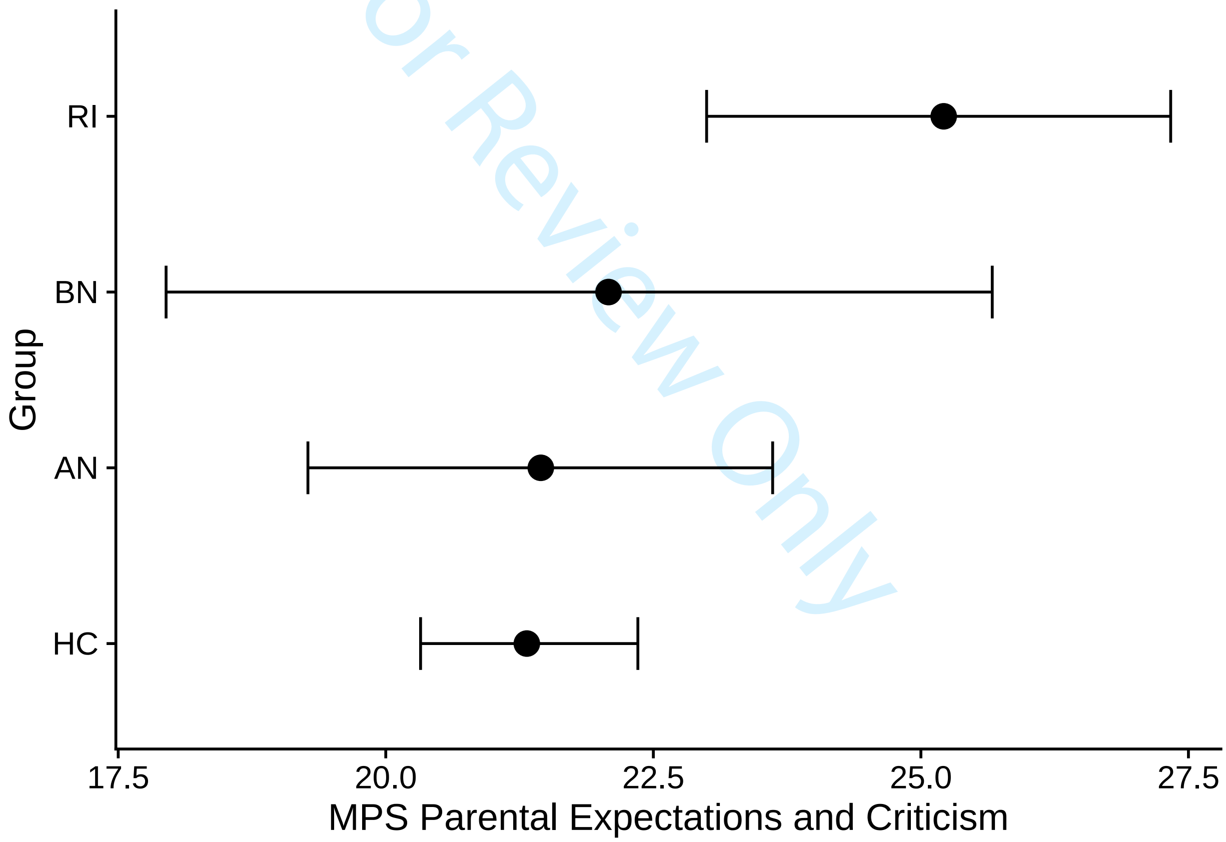


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	21.324	0.520	20.291	22.310
groupAN	0.125	1.172	-2.142	2.459
groupBN	0.695	1.989	-3.209	4.637
groupRI	3.875	1.202	1.431	6.178

Predicted effect of group on the MPS Parental Expectations and Criticism



Interpretation

The 95% credibility intervals for the difference in the MPS-PEC scores between the HC and the RI group do not include zero, indicating credible differences in Parental Expectations and Criticism between these two groups. We found no credible difference in MPS-PEC between the HC group, on the one side, and the AN and BN groups, on the other.

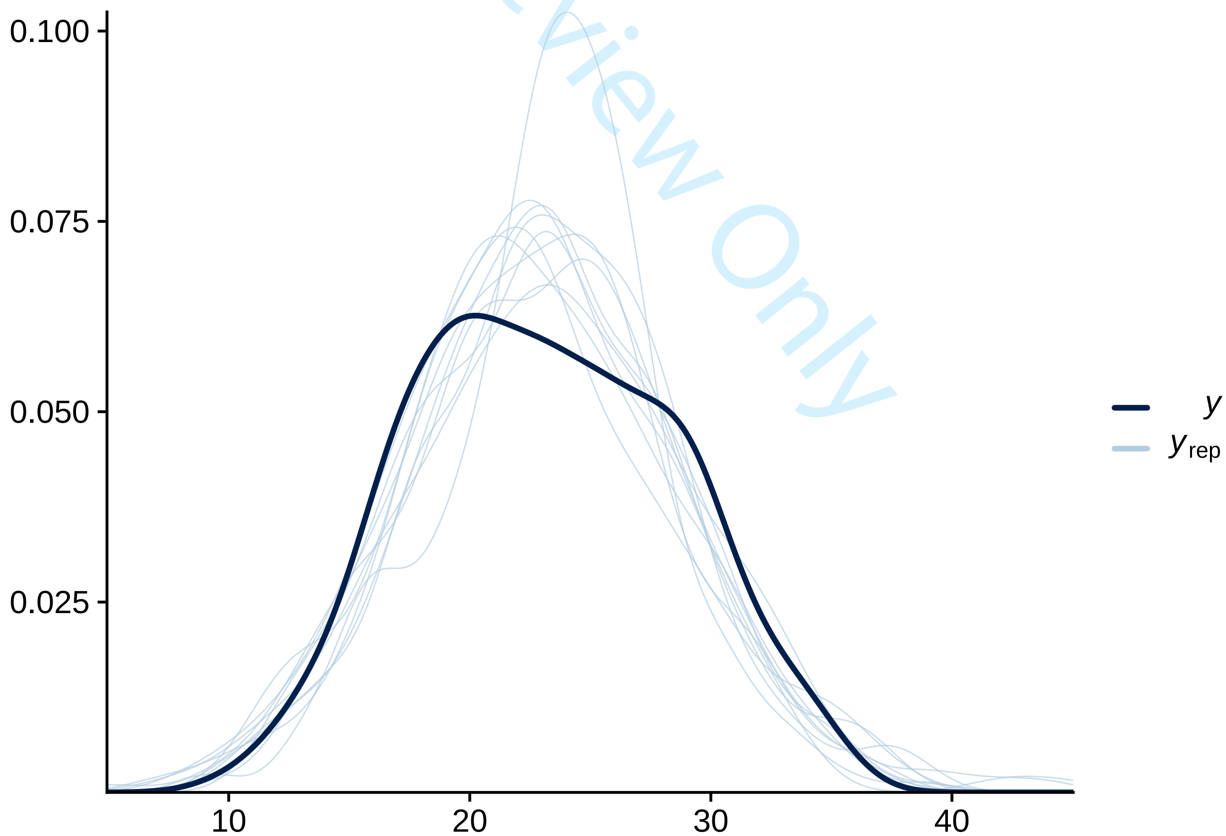
Personal Standards

We used a Bayesian regression model to examine the MPS PS score differences among groups.

```
m7_ps <- brm(
  mps_ps ~ group,
  data = mps_df,
  family = gaussian(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

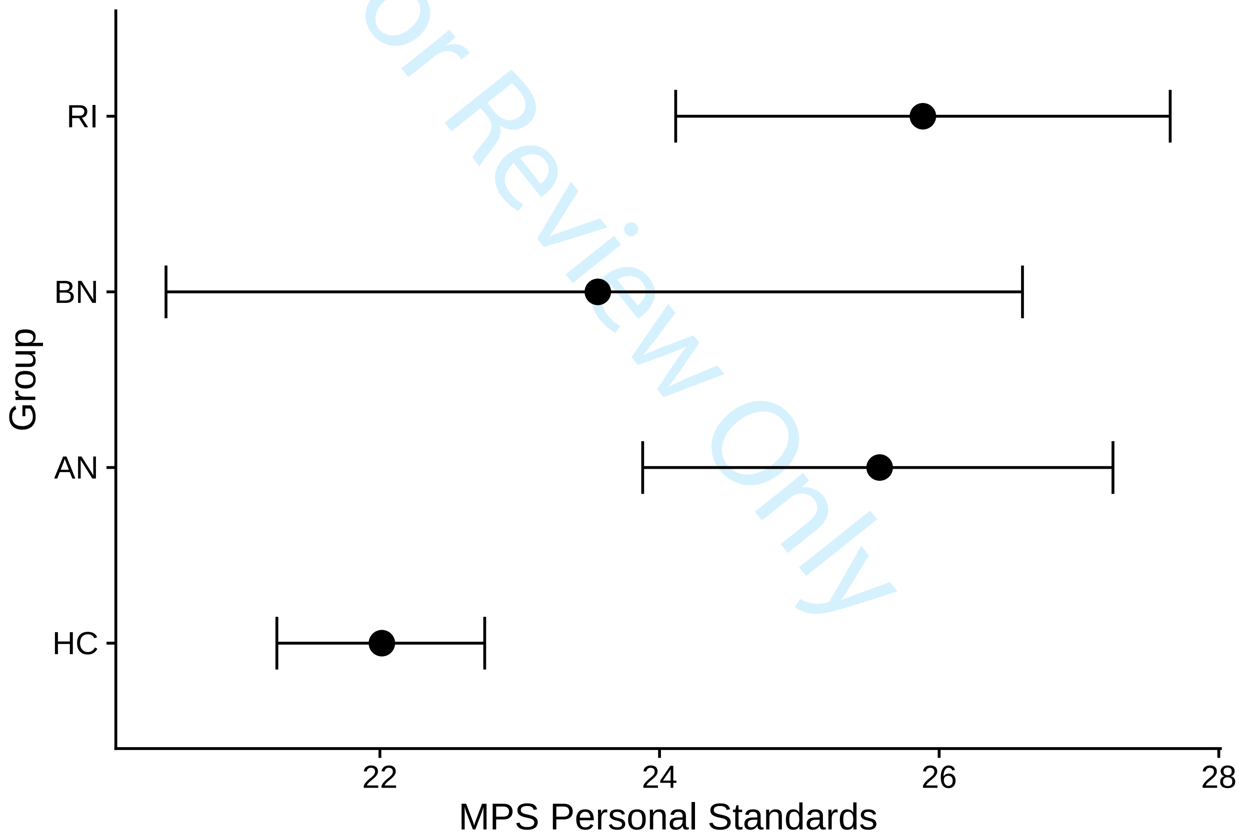


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	22.014	0.382	21.259	22.741
groupAN	3.559	0.941	1.714	5.413
groupBN	1.530	1.614	-1.597	4.699
groupRI	3.872	0.978	1.874	5.759

Predicted effect of group on the MPS Personal Standards scores



Interpretation

The 95% credibility intervals for the difference in the MPS-PS scores between the HC group, on the one side, and the AN and RI groups, on the other, do not include zero, indicating credible differences in Personal standards between these groups. We found no credible difference in MPS-PS between the HC group and the BN groups.

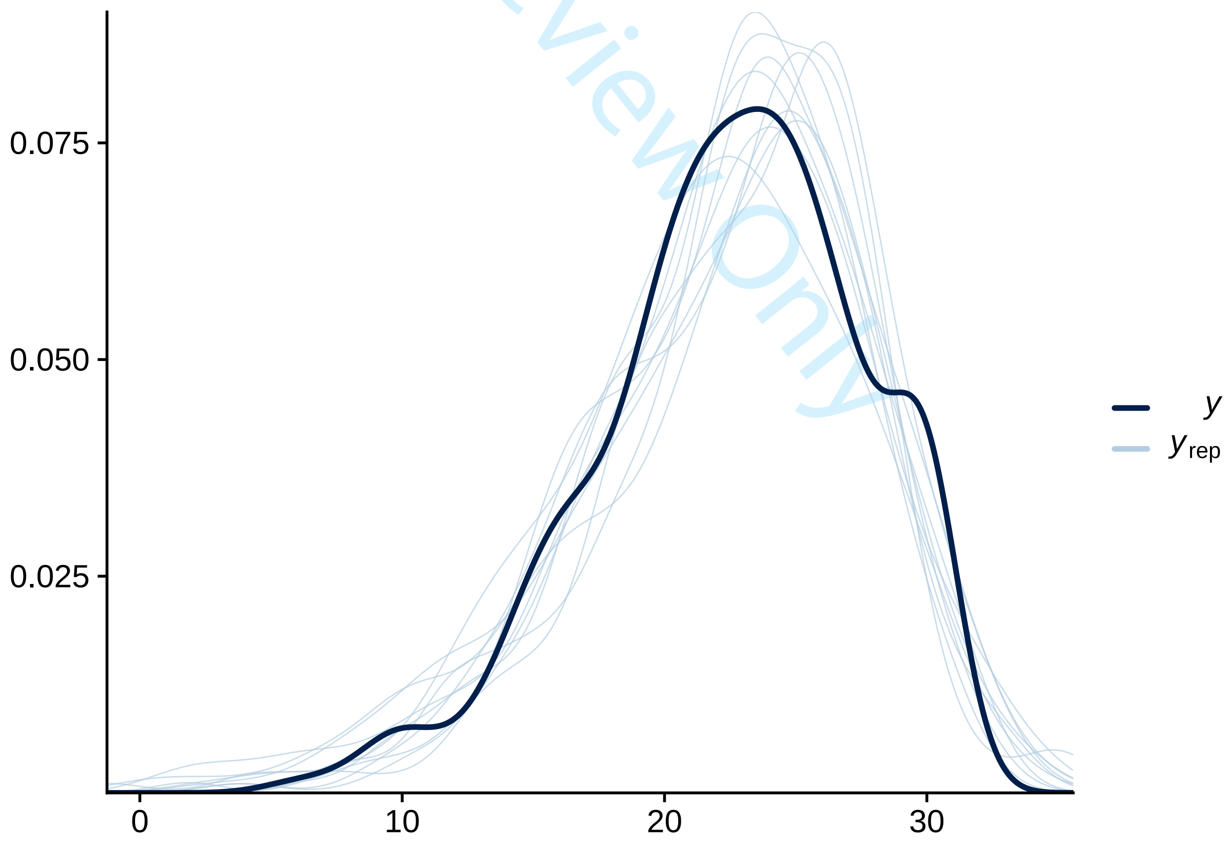
Organization

We used a Bayesian regression model to examine the MPS-O score differences among groups.

```
m7_or <- brm(
  mps_or ~ group,
  data = mps_df,
  family = skew_normal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```


Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```

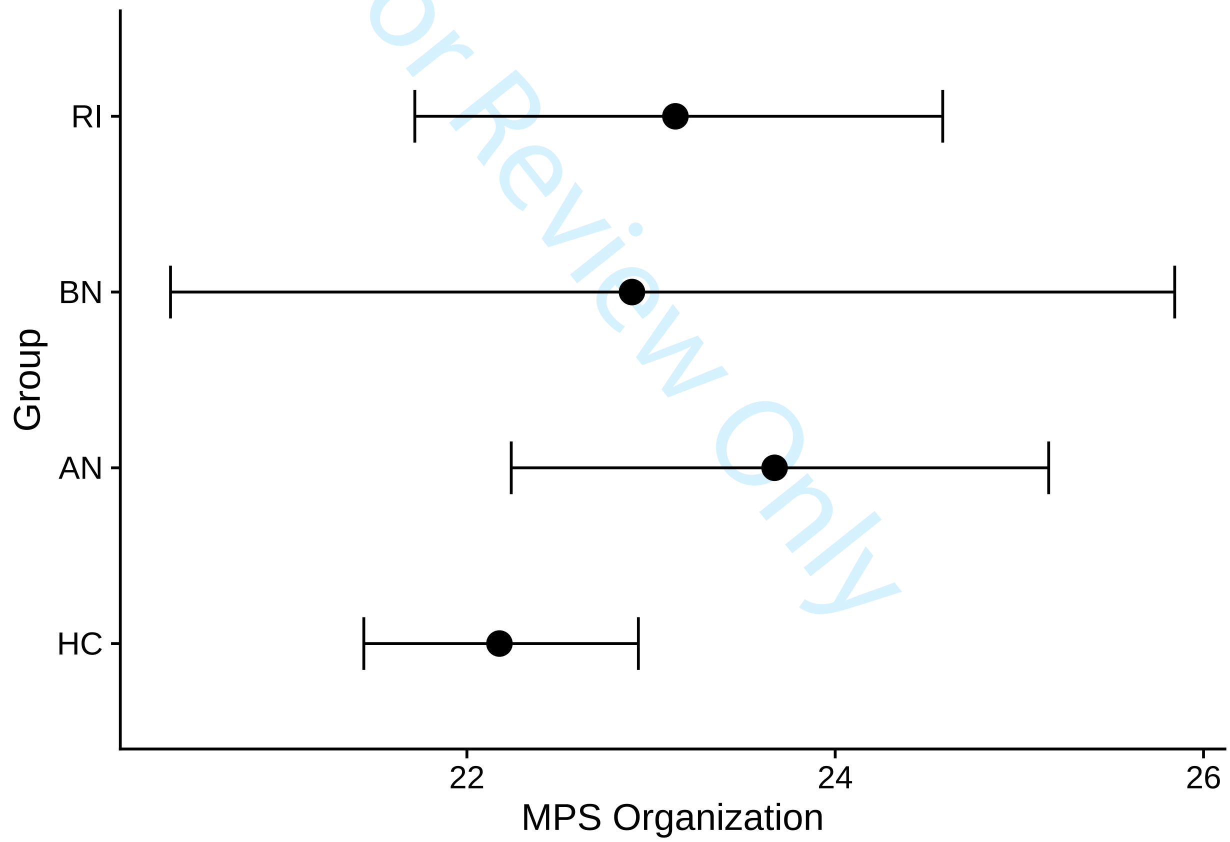


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	22.181	0.383	21.443	22.932
groupAN	1.504	0.811	-0.046	3.090
groupBN	0.770	1.414	-1.918	3.629
groupRI	0.955	0.808	-0.535	2.583

Predicted effect of group on the MPS Organization scores



Interpretation

The 95% credibility intervals for the difference in the MPS-O scores between the HC group, on the one side, and the AN, BN, and RI groups, on the other, do include zero, indicating that there are no credible differences in Organization between these groups.

To summarize, our findings are in line with the notion that maladaptive perfectionism is a factor linked to both AN and BN (Dahlenburg et al., 2019).

Eating Attitude Test-26 (EAT-26)

The average scores on the Dieting, Bulimia and Food Preoccupation, and Oral Control scales, together with their standard deviations, are shown below as a function of group. The tables also show the EAT-26 total scores.

```
## distinct: removed 219 rows (46%), 259 rows remaining
## group_by: one grouping variable (group)
## summarize: now 4 rows and 5 columns, ungrouped
```

group	EAT26_dieting	EAT26_bulimia	EAT26_oralcontrol	EAT26_tot
HC	2.47	0.57	1.10	4.14
AN	19.67	7.06	9.14	35.86
BN	19.09	7.64	5.64	32.36
RI	15.51	6.46	3.63	25.60

```
## group_by: one grouping variable (group)
## summarize: now 4 rows and 5 columns, ungrouped
```

group	SD_dieting	SD_bulimia	SD_oralcontrol	SD_tot
HC	3.25	1.31	1.81	4.45
AN	10.60	3.70	5.91	17.96
BN	13.60	4.78	7.49	23.15
RI	7.85	3.81	3.83	12.45

Dieting

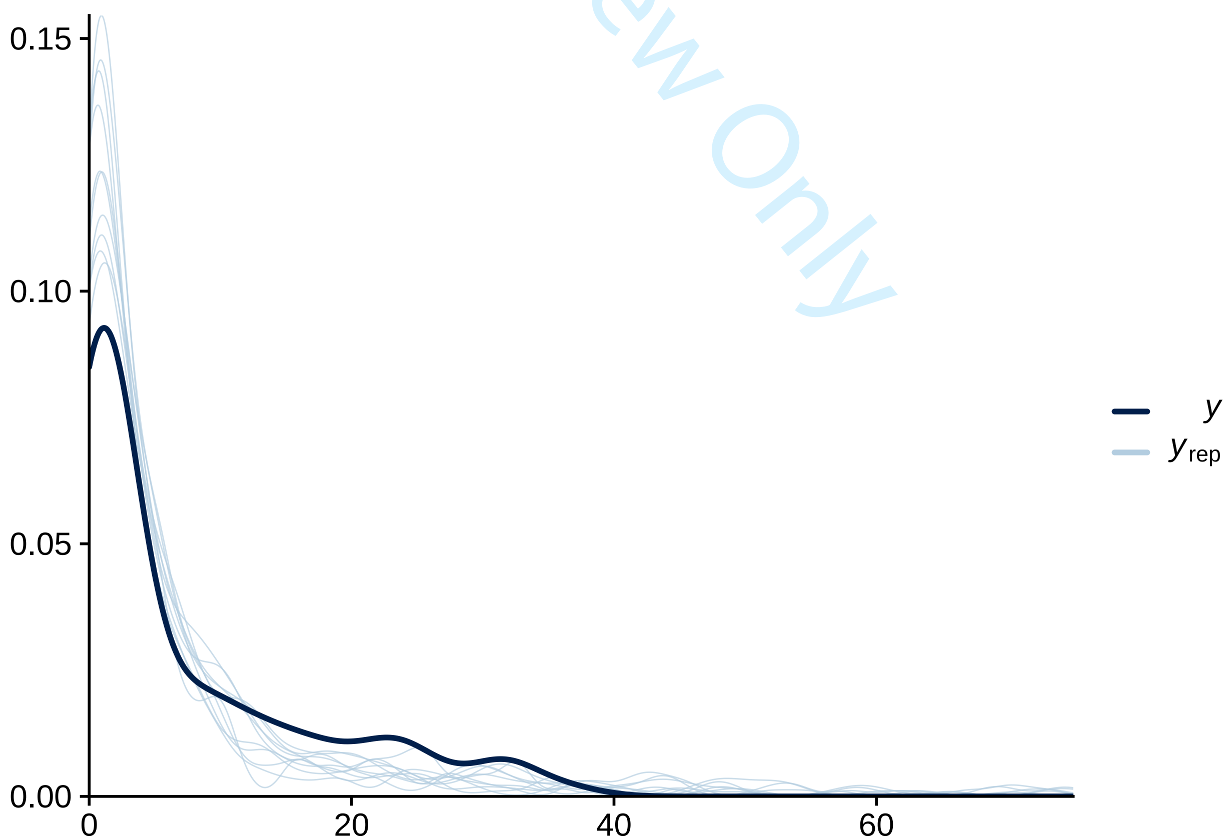
We used a Bayesian regression model to examine the EAT-26 Dieting score differences among groups.

```
m8_d <- brm(
  dieting ~ group,
  data = eat26_df,
  family = hurdle_lognormal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.

## Warning: Removed 12 rows containing non-finite values (`stat_density()`).
```

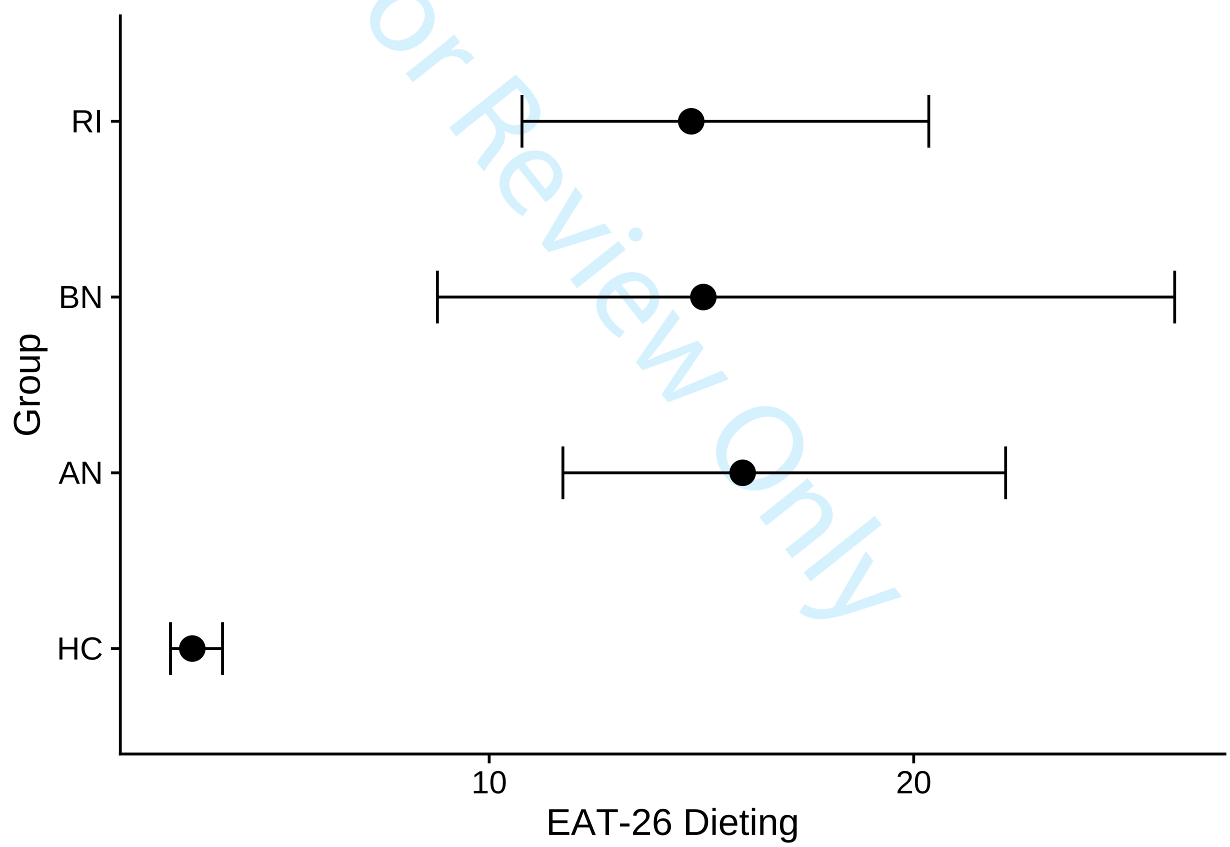


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	0.979	0.085	0.812	1.145
groupAN	1.668	0.170	1.352	2.011
groupBN	1.611	0.285	1.051	2.168
groupRI	1.589	0.174	1.242	1.922

Predicted effect of group on the EAT-26 Dieting scores



Interpretation

The 95% credibility intervals for the difference in the EAT-26 Dieting scores between the HC group, on the one side, and the AN, BN, and RI groups, on the other, do not include zero, indicating a credible elevation in the EAT-26 Dieting scores of the AN, BN, and RI groups with respect to the HC group.

Bulimia and Food Preoccupation

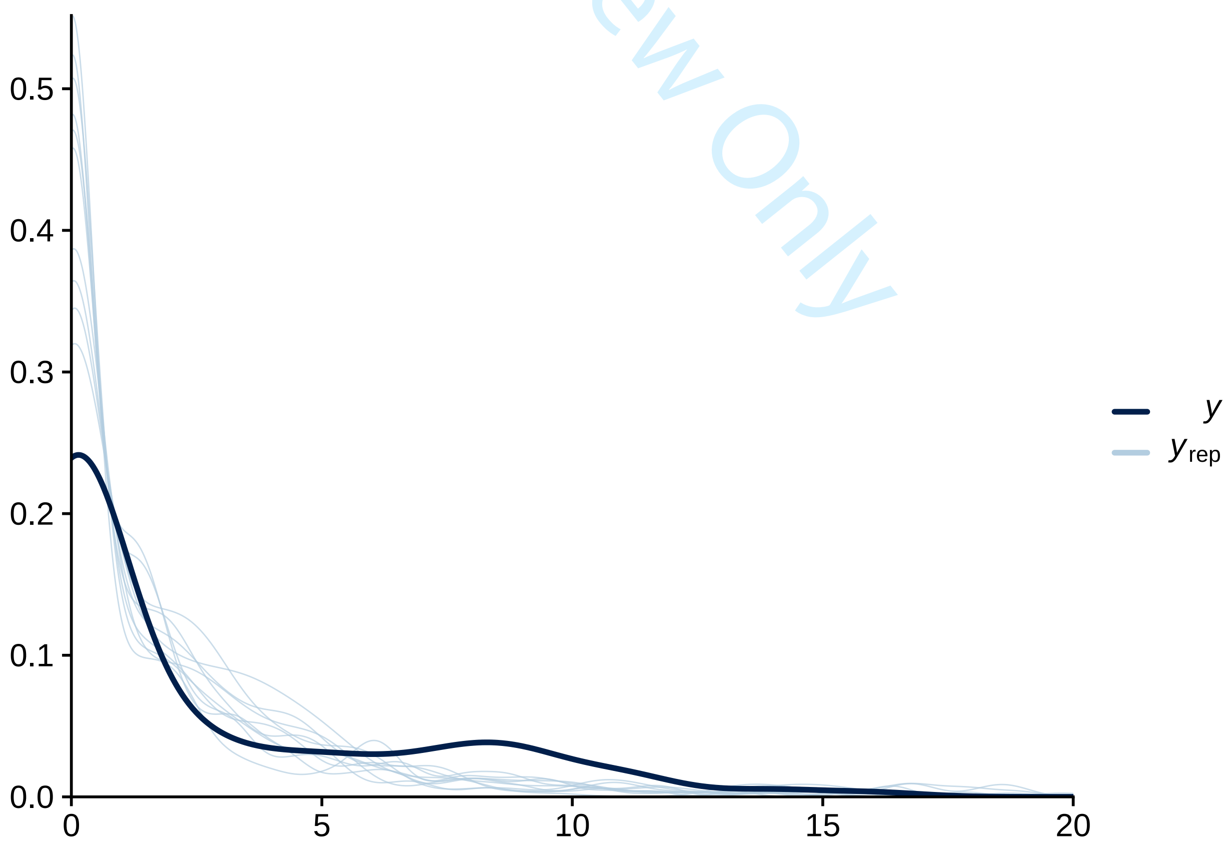
We used a Bayesian regression model to examine the EAT-26 Bulimia and Food Preoccupation score differences among groups.

```
m8_b <- brm(
  bulimia ~ group,
  data = eat26_df,
  family = hurdle_lognormal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.

## Warning: Removed 7 rows containing non-finite values (`stat_density()`).
```

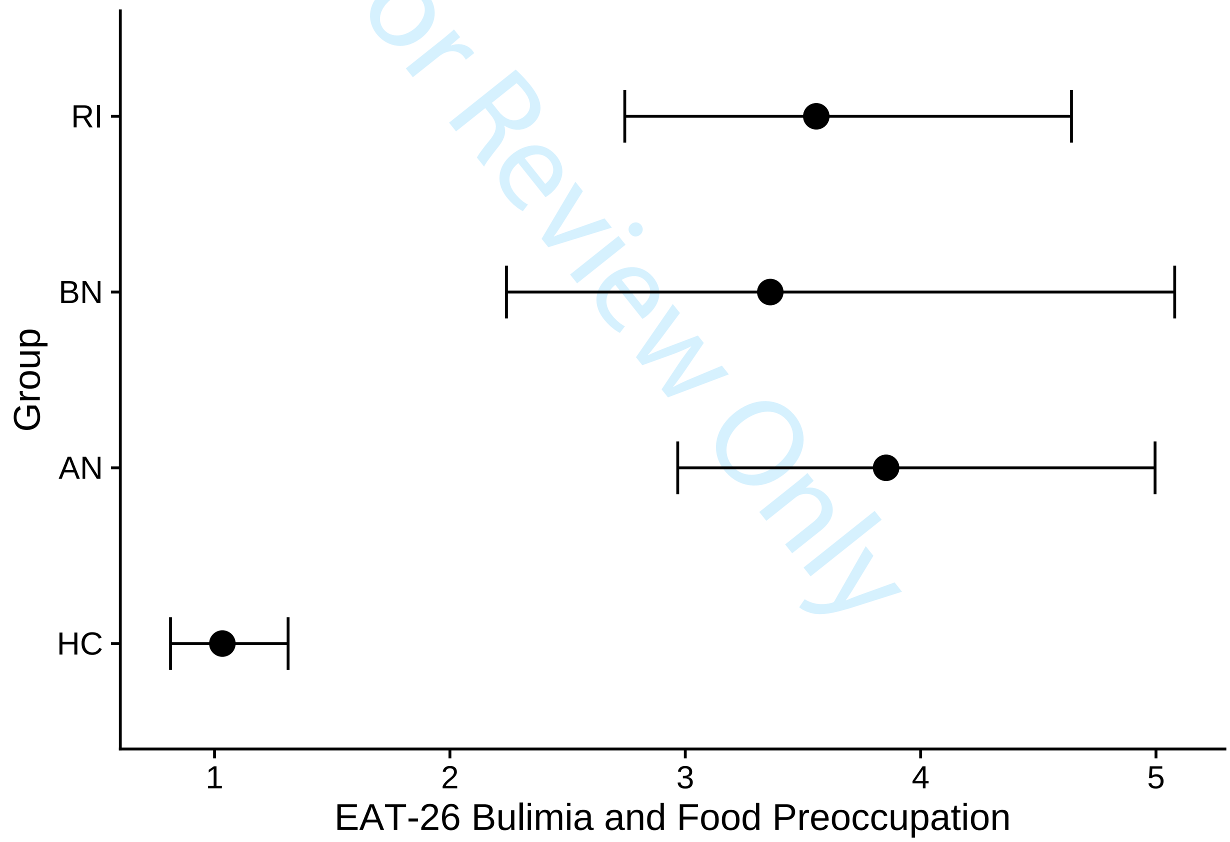


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	0.587	0.097	0.404	0.782
groupAN	1.315	0.147	1.027	1.602
groupBN	1.183	0.218	0.734	1.587
groupRI	1.236	0.149	0.961	1.530

Predicted effect of group on the EAT-26 Dieting scores



Interpretation

The 95% credibility intervals for the difference in the EAT-26 Bulimia and Food Preoccupation scores between the HC group, on the one side, and the AN, BN, and RI groups, on the other, do not include zero, indicating a credible elevation in the EAT-26 Bulimia and Food Preoccupation scores of the clinical and at-risk groups with respect to the HC group.

Oral Control

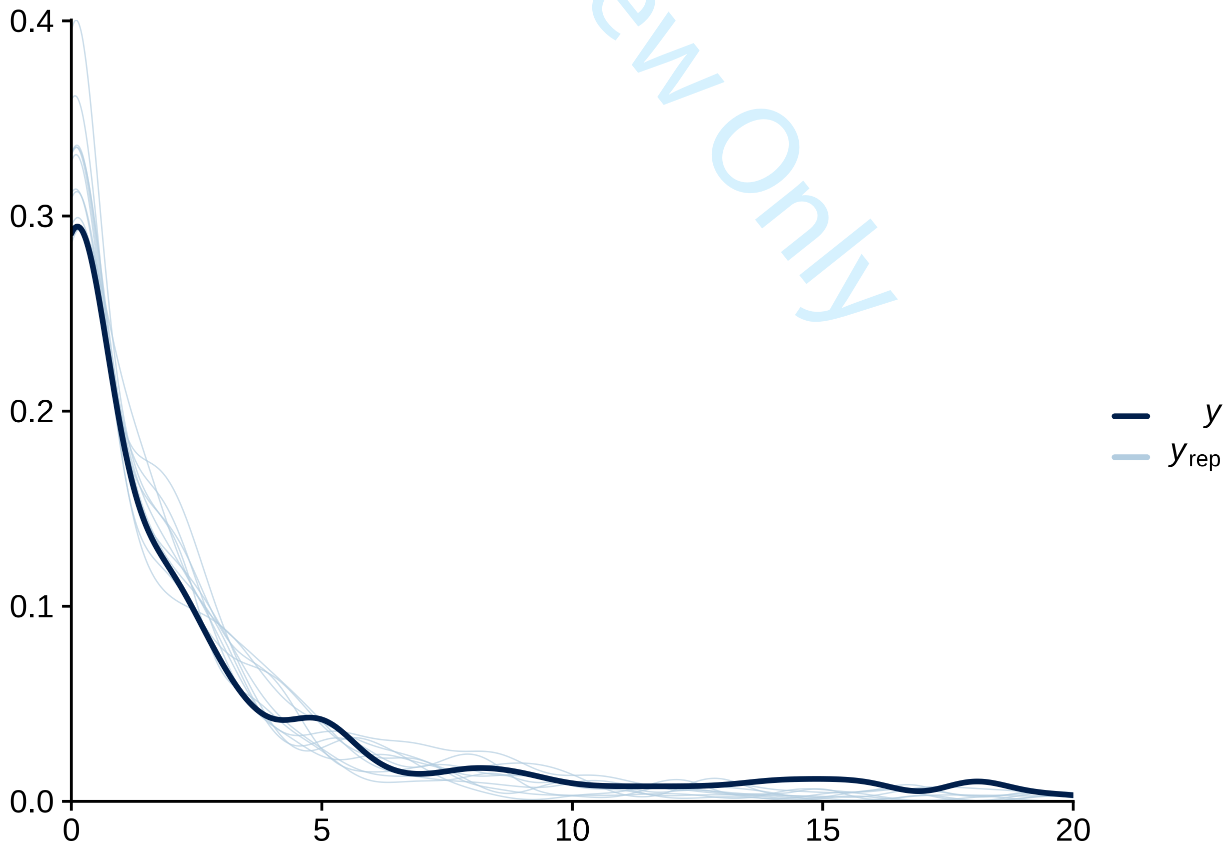
We used a Bayesian regression model to examine the EAT-26 Oral Control score differences among groups.

```
m8_oc <- brm(
  oral_control ~ group,
  data = eat26_df,
  family = hurdle_lognormal(),
  iter = 4000,
  cores = parallel::detectCores(),
  backend = "cmdstan",
  refresh = 0,
  silent = TRUE
)
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.

## Warning: Removed 26 rows containing non-finite values (`stat_density()`).
```

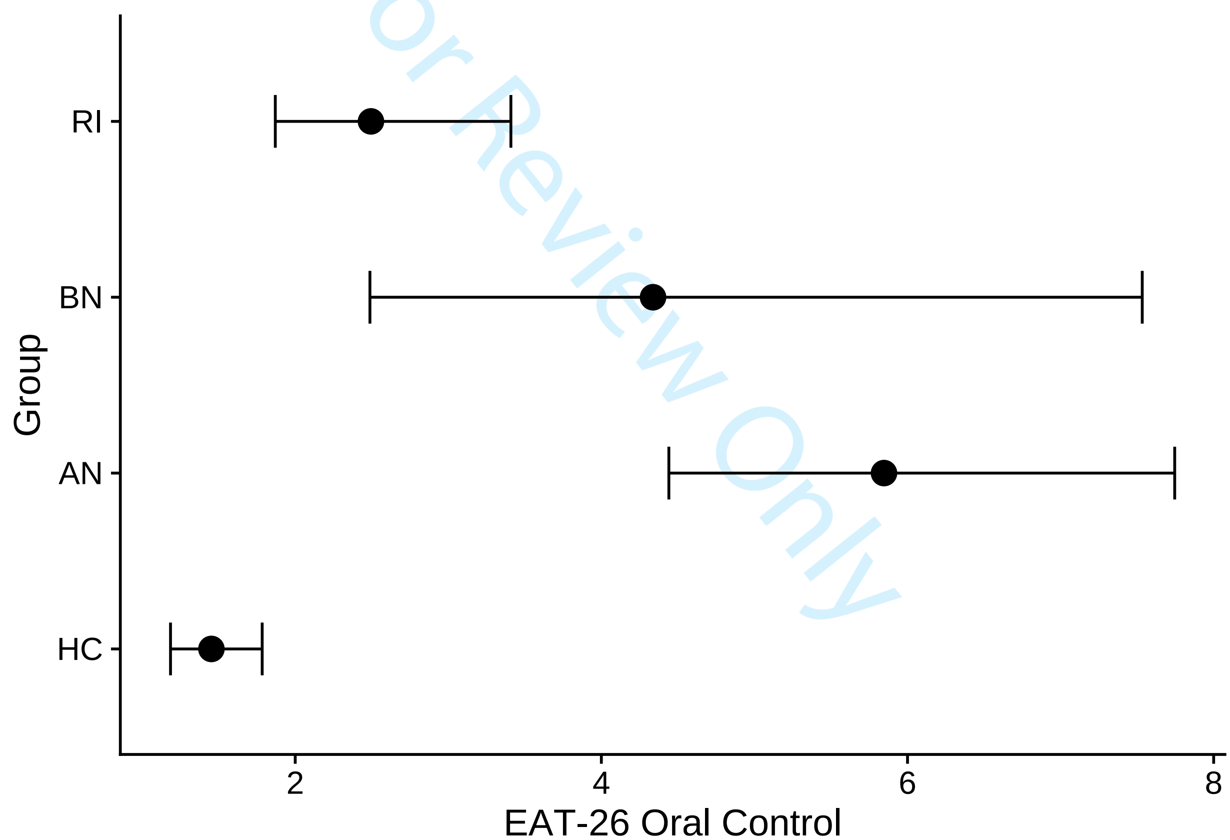


Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	0.689	0.084	0.526	0.852
groupAN	1.392	0.152	1.095	1.688
groupBN	1.093	0.290	0.505	1.641
groupRI	0.542	0.161	0.224	0.849

Predicted effect of group on the EAT-26 Oral Control



Interpretation

The previous figure show the 95% credibility intervals for the difference in the EAT-26 Oral Control scores between the HC group and the AN, BN, and RI groups. These intervals do not contain the null value of zero, which suggests that there is a credible elevation in the EAT-26 Oral Control scores of the clinical and at-risk groups in comparison to the HC group.

Together, the results of the previous statistical analyses provide confirmatory evidence that the EAT-26, a widely employed tool for assessing eating disorder risk, not only discriminates between healthy controls and AN and BN patients, but also between healthy controls and individuals at risk of developing eating disorders.

Procedure

Participants' height was determined using a wall-mounted stadiometer, and their weight was measured using a beam balance scale. To evaluate general cognitive functioning, participants completed Raven's Standard Progressive Matrices test (Jean Raven, 2003). Following the cognitive and psychological-clinical evaluations, participants underwent a series of psychometric assessments and completed the PRL task during two separate sessions, which were conducted over a period not exceeding three days. The psychologists responsible for administering the questionnaires and behavioral tasks to both the patient and control groups were blinded to the objectives of the investigation.

Probabilistic Reversal Learning Task

The PRL task consisted of two blocks. The first block aimed to test the domain-specificity hypothesis by pairing food-related images, such as a piece of cake, with food-unrelated images, such as a lamp. In contrast, the second block served as a control task and solely employed food-unrelated images.

Each trial in the food-related image block presented two images randomly selected from a set of food-related and food-unrelated categories. All images used in the study were obtained from the International Affective Picture System (IAPS) database (Lang et al., 2005). The food-related category included images of french fries, cake, pancake, cheeseburger, and cupcake (IAPS #7461, 7260, 7470, 7451, 7405) while the food-unrelated category comprised images of a lamp, book, umbrella, basket, and clothespin (IAPS #7175, 7090, 7150, 7041, 7052). For the control task, five images were utilized for each of the two food-unrelated categories of images, i.e., five images of flowers (IAPS #5000, 5001, 5020, 5030, 5202) and five images of objects (IAPS #7010, 7020, 7034, 7056, 7170).

The Psychtoolbox extensions in MATLAB (MathWorks) were used to program the tasks (Brainard, 1997), as described in the methods section of the main text. Trials were presented in an interleaved fashion, with an inter-trial interval randomly drawn from a uniform distribution (uniform 0.5-1.5 s). Stimuli were presented until the participant responded or until the time limit of 2.5 seconds was reached. Participants completed a training block of 20 trials prior to the experiment.

Reinforcement learning drift diffusion model (RLDDM)

We used a reinforcement learning drift diffusion model [RLDDM; Pedersen & Frank (2020)] to investigate the impact of illness-related information, which was irrelevant to the outcome, on decision-making. The RLDDM consists of two key components: one describes how reward feedback is employed to update value expectations and the other describes how an agent uses these expectations to arrive at a decision.

δ-learning rule and decision making

The first component characterizes the learning process in terms of the delta learning rule (Rescorla, 1972):

$$Q_{a,i} = Q_{a,i-1} + \alpha(I_{a,i-1} - Q_{a,i-1}).$$

Here, Q refers to the expected values for option a on trial i, I represents the reward (with values 1 or 0), and α is the leaning rate, which scales the difference between the expected and actual rewards.

The second component describes the selection rule for reinforced options. Typically, a softmax function is used, where the probability of selecting option a depends on its expected value relative to other options n, scaled by the inverse temperature parameter β:

$$p_{a,i} = \frac{e^{\beta Q_{a,i}}}{\sum_{j=1}^{n} e^{\beta Q_{j,i}}}.$$

In the RLDDM, instead, this second component of decision-making is replaced by a Drift-Diffusion Model [DDM; Ratcliff & McKoon (2008)] which assumes a stochastic accumulation of evidence on each trial, and includes four parameters: A drift rate parameter (v), a decision threshold parameter (a), a non-decision time parameter (t), and a starting point parameter (z).

RLDDM parameters

The RLDDM consists of six fundamental parameters: the positive learning rate (α^+), negative learning rate (α^-), drift rate (v), decision threshold (a), non-decision time (t), and starting point bias (z) parameters.

- The parameter α reflects the learning rate in the Rescorla-Wagner δ-learning rule (Rescorla, 1972), whereby a higher learning rate leads to rapid adaptation to reward expectations, while a lower learning rate results in slower adaptation. The α^+ and α^- parameters capture the impact of reinforcements and punishments, respectively.
- The drift rate ν reflects the average speed of evidence accumulation toward one decision.
- The decision threshold a determines the distance between two decision boundaries, with an increase of a resulting in a slower but more accurate decision and a decrease of a leading to a faster but more error-prone decision.
- The non-decision time t captures the time spent on stimuli encoding or motor execution, which is not used for evidence accumulation.
- The starting point parameter z quantifies a potential initial bias toward one or the other boundary in the absence of any stimulus evidence.

Estimation

The posterior distribution of group and individual parameters of the RLHDDM model were estimated in a hierarchical Bayesian framework using the "HDDMrl" module of the "HDDM" (version 0.9.8) Python package (for a detailed description of the model, see Fengler et al., 2021; Pedersen & Frank, 2020; Wiecki et al., 2013).

Priors

The Bayesian posterior estimations of the RLDDM rely on informative priors obtained from a prior meta-analysis (Wiecki et al., 2013) for the DDM aspect of the model. On the other hand, non-informative broad normal distributions, centered at 0.5 after transformation, are used for the learning rate parameters (positive and negative).

Data analysis

Quality control

To ensure data quality, participants who performed below chance level in the PRL task were excluded from further analysis (e.g., Geisler et al., 2017). A total of 278 participants met the quality control criterion and were included in subsequent analyses. This sample comprised 37 individuals with AN, 12 individuals with BN, 198 HCs, and 31 healthy controls at risk of developing eating disorders (RI).

Models' comparison

The study utilized a minimal model as the basis for constructing new models, wherein all parameters were permitted to vary by condition (i.e., food-related vs. food-unrelated information) and group. This decision was based on the absence of a priori knowledge suggesting that outcome-irrelevant effects were limited to a particular parameter of the RLDDM or that they varied across groups. Markov Chain Monte Carlo (MCMC) sampling was employed to estimate these models, with 2000 traces being sampled following a burn-in period of 500. The Deviance Information Criterion (DIC) was computed for each model. We selected the model with the model with the lowest DIC.

The winning model was subsequently better estimated using 13000 traces and a 3000 burn-in (Kruschke, 2014).

Collinearity check

As shown in the following figures, for all four groups the correlation between the parameters is generally low.

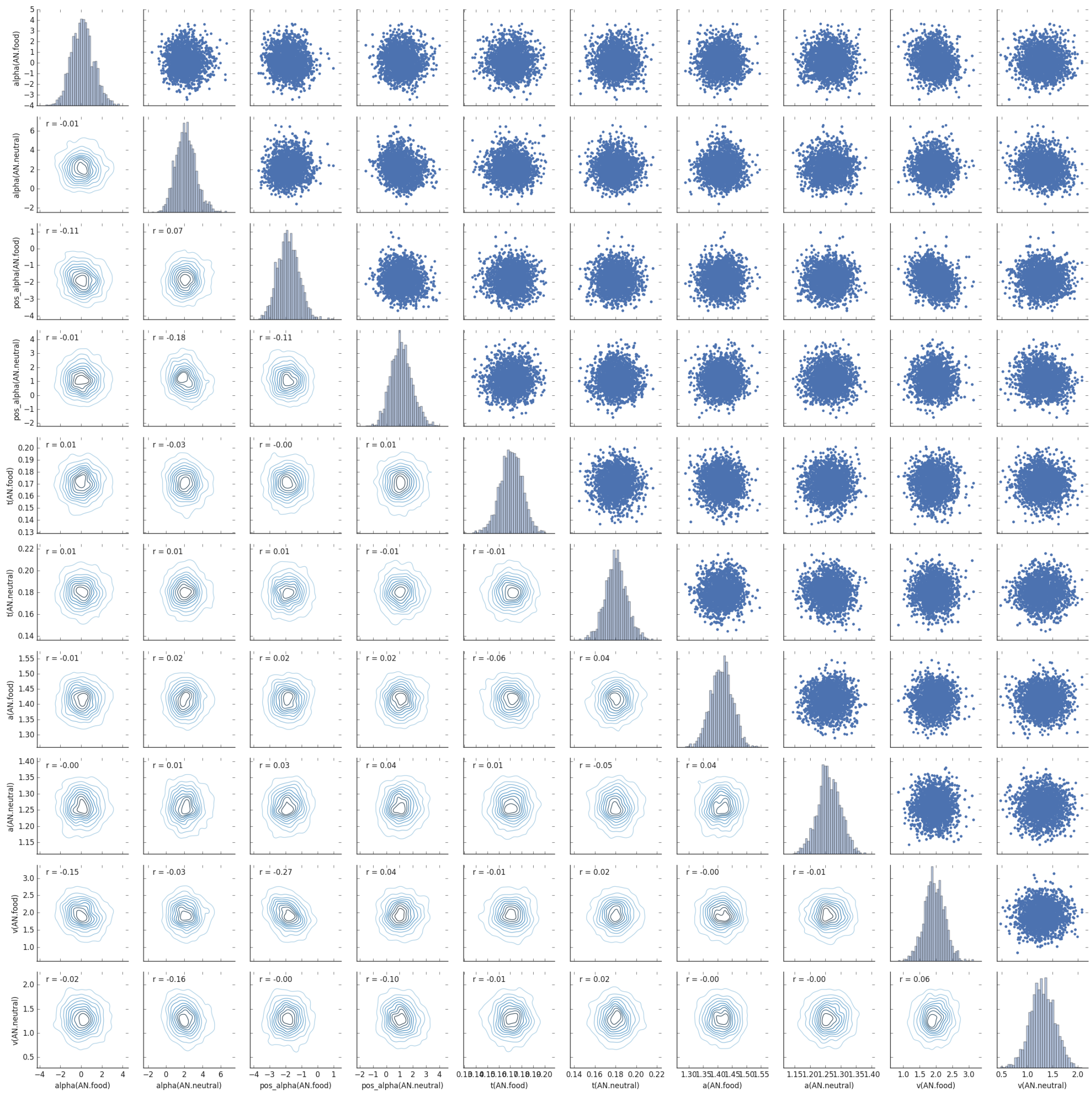


Figure 1: Joint posterior distribution of RLDMM parameters: AN group.

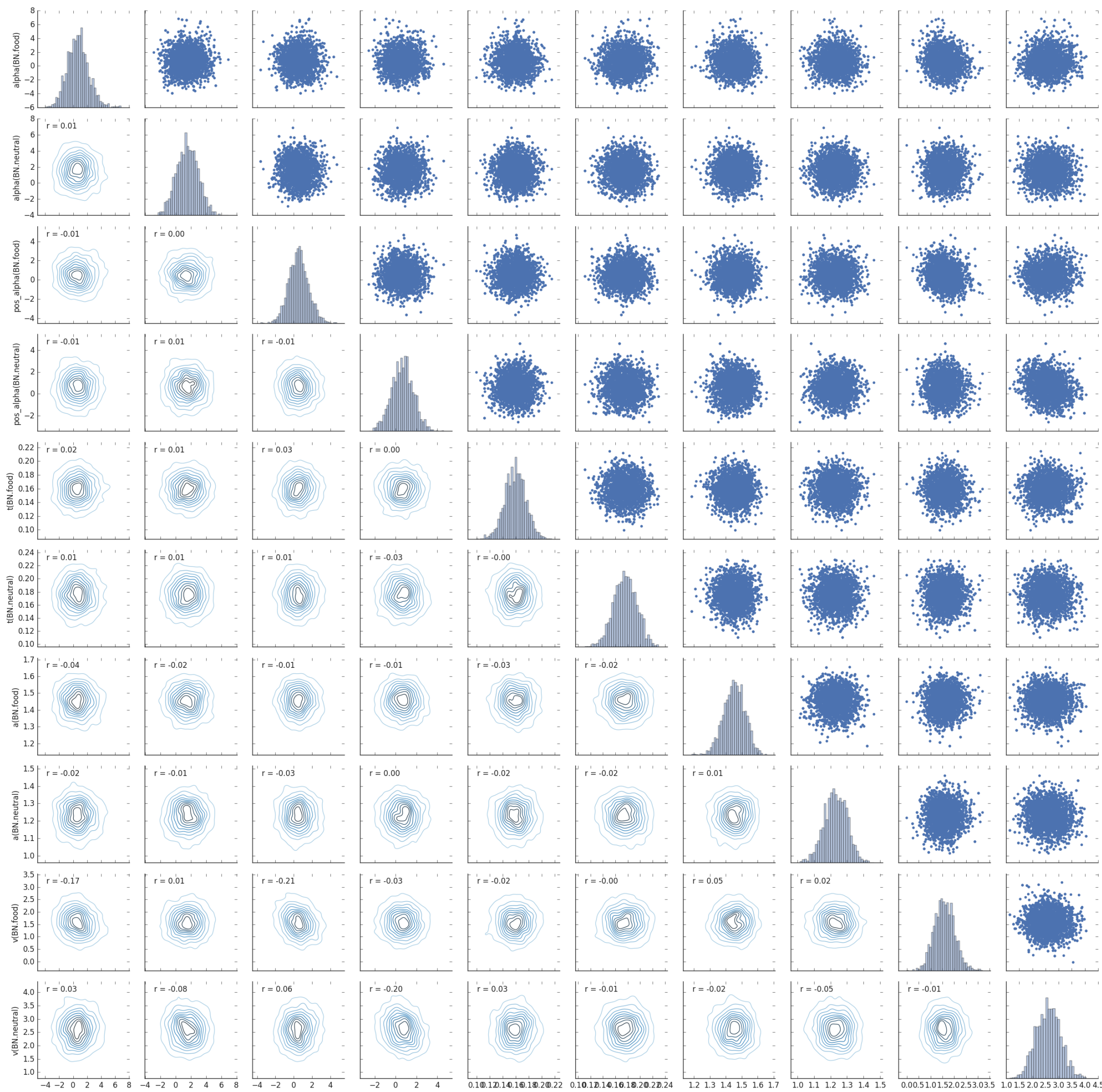


Figure 2: Joint posterior distribution of RLDMM parameters: BN group.

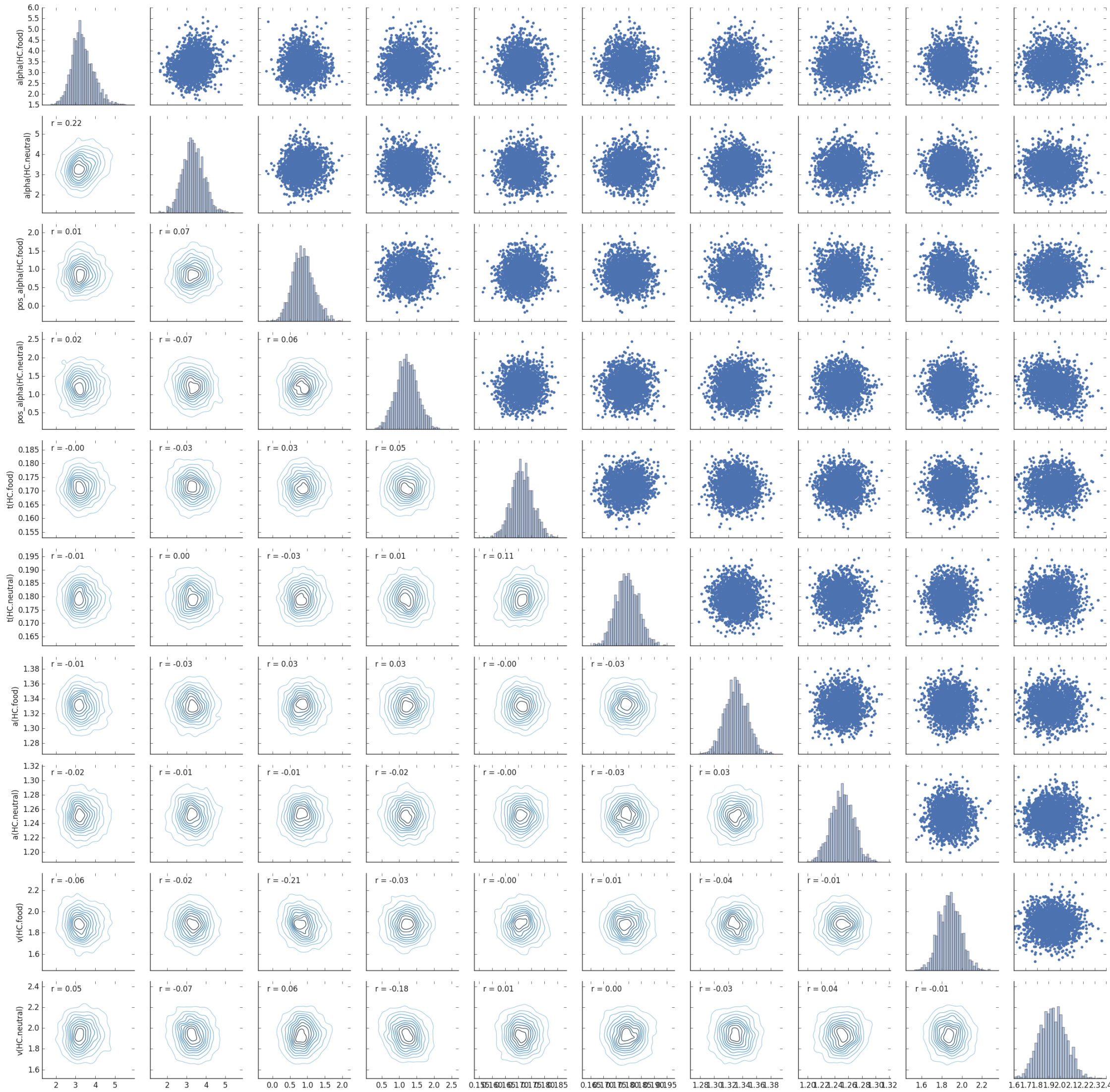


Figure 3: Joint posterior distribution of RLDMM parameters: HC group.

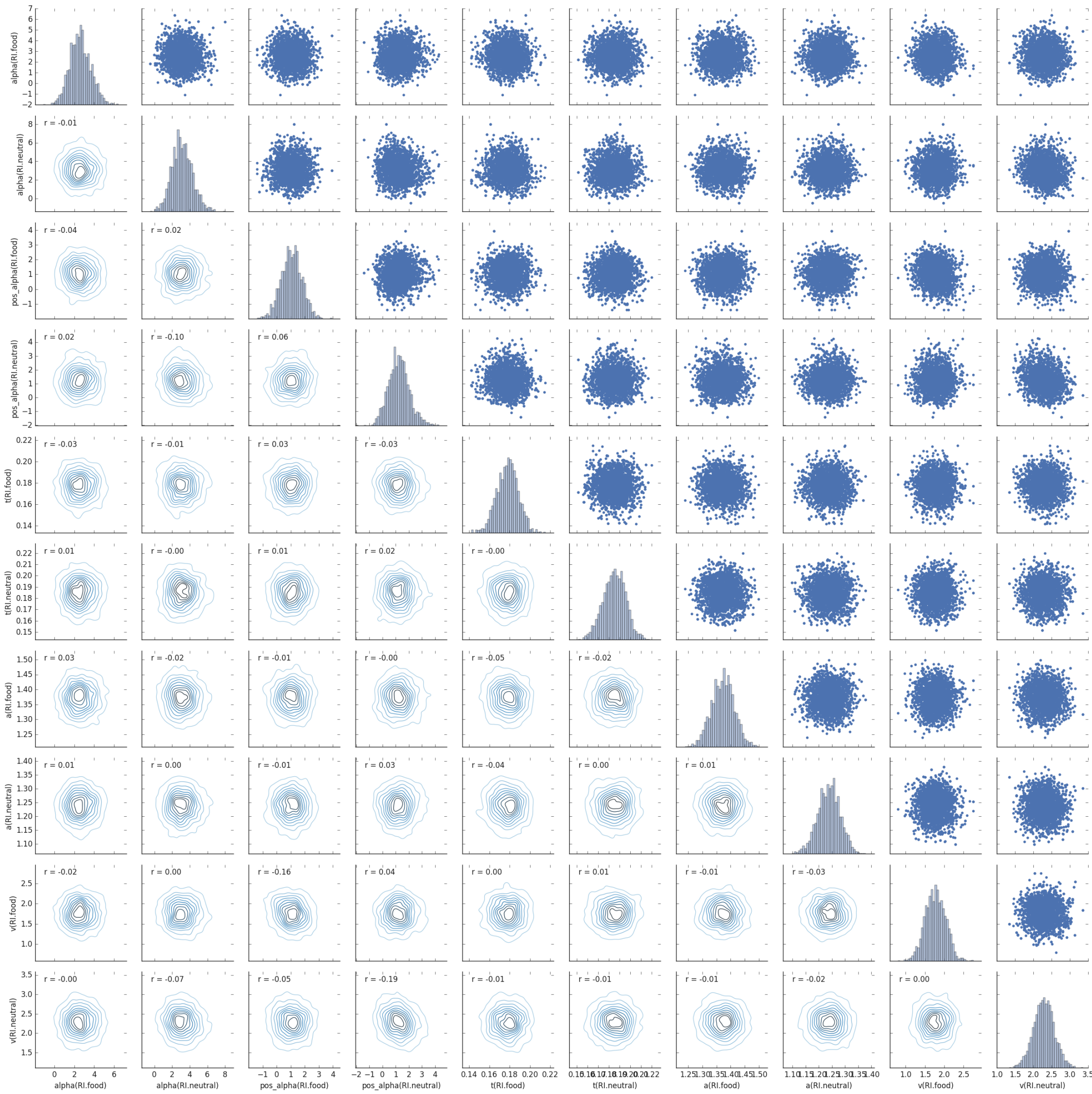


Figure 4: Joint posterior distribution of RLDMM parameters: RI group.

Posterior predictive checks

To assess model validity, we performed posterior predictive checks. This involved simulating data using estimated parameters and comparing observed and simulated results. We generated the simulated dataset by repeating the simulation process 500 times for each subject in a sample dataset.

PPC for learning rate To evaluate the choice proportion for the best option across learning in both observed and simulated data, we binned the trials and plotted the 90% highest density intervals of the mean responses. The following figures illustrate the rate of selecting the best option during learning. The 90% highest density interval of the means across simulated datasets captures the uncertainty present in the generated data.

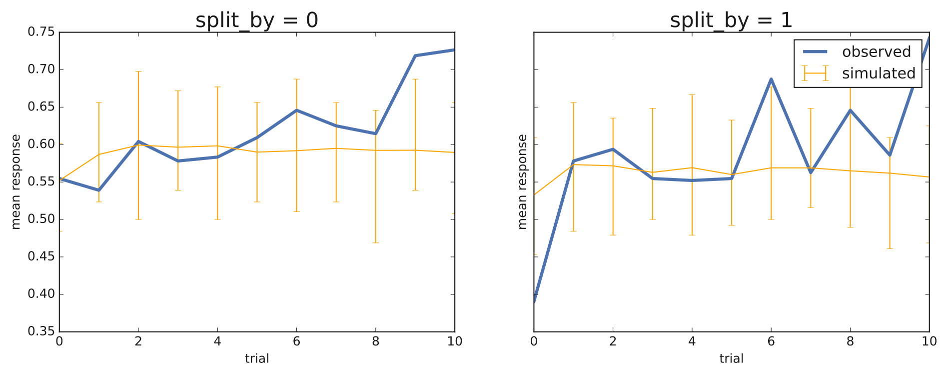


Figure 5: Observed and predicted learning rates across conditions: AN group.

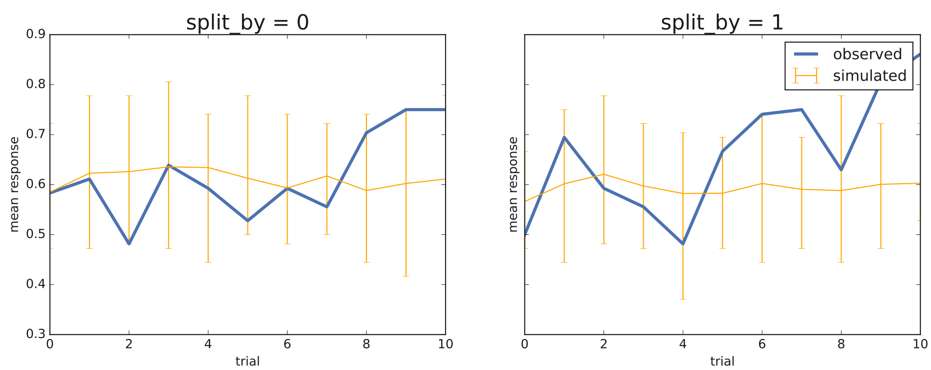


Figure 6: Observed and predicted learning rates across conditions: BN group.

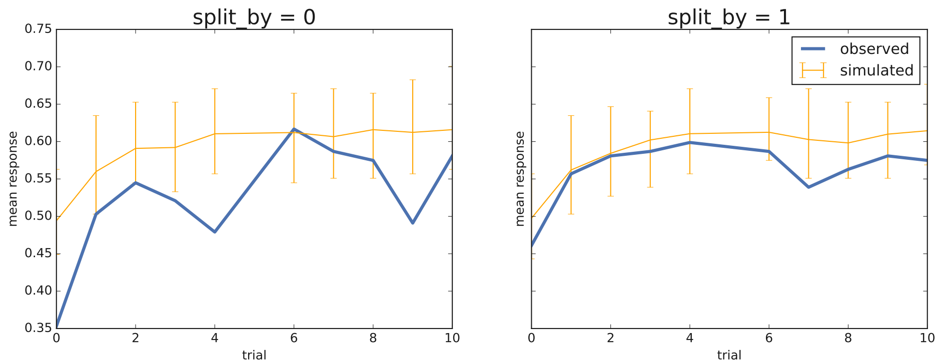


Figure 7: Observed and predicted learning rates across conditions: HC group.

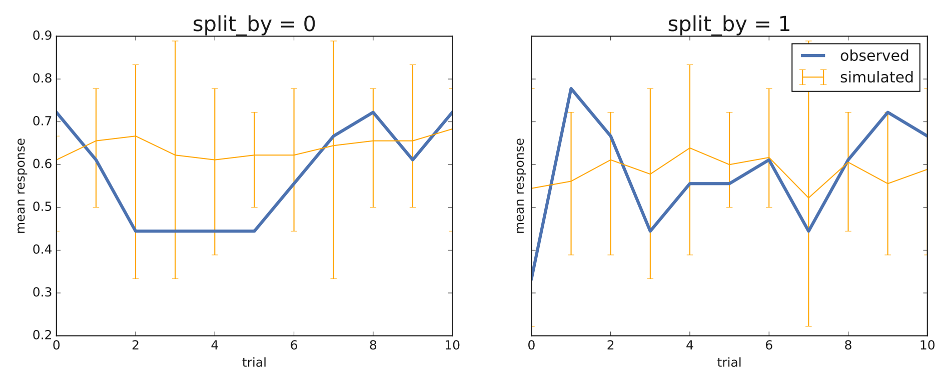


Figure 8: Observed and predicted learning rates across conditions: RI group.

RT The density plots of observed and predicted reaction time across conditions are presented in the following figures. To distinguish upper and lower bound responses, reaction times for lower boundary choices (i.e., worst option choices) were set to be negative (0-RT). There is a good agreement between the observed and predicted values.

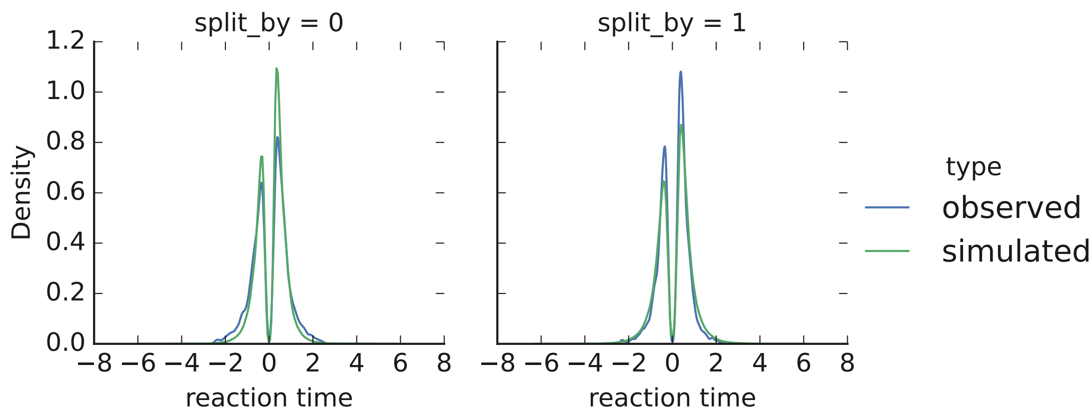


Figure 9: Observed and predicted reaction time across conditions: AN group.

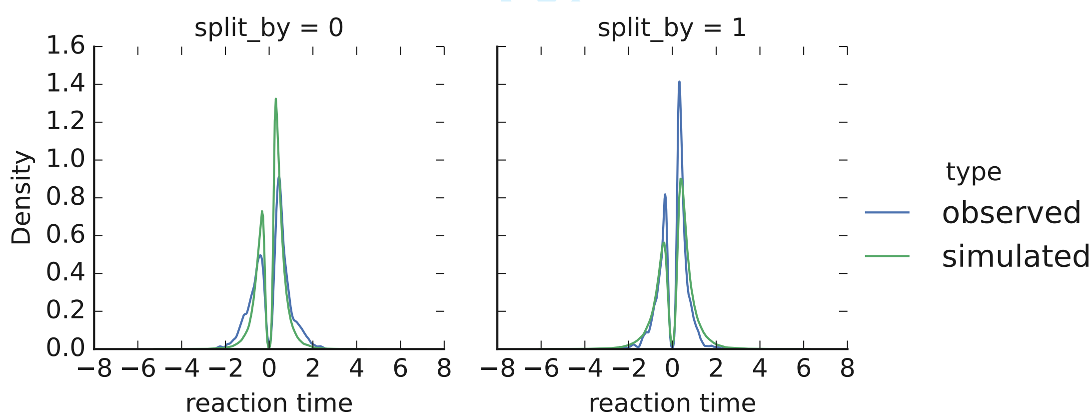


Figure 10: Observed and predicted reaction time across conditions: BN group.

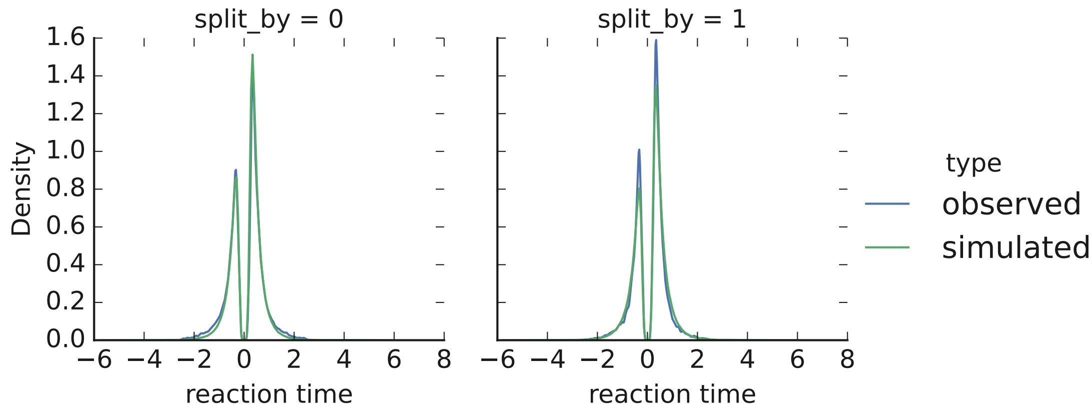


Figure 11: Observed and predicted reaction time across conditions: HC group.

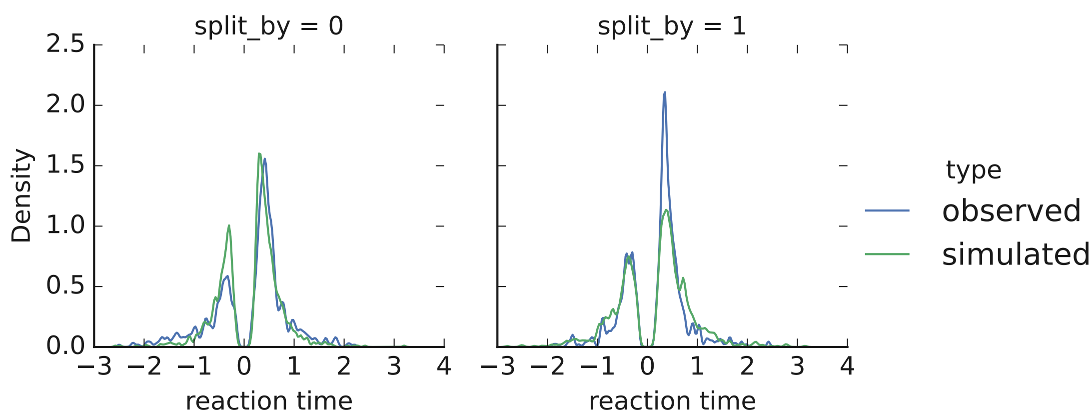


Figure 12: Observed and predicted reaction time across conditions: RI group.

Drift-diffusion process

Drift rate

Consistent with the hypothesis that contextual factors affect cognitive processing was also a result pertaining to the evidence accumulation rate in the drift-diffusion process (parameter v of the RLDDM). Specifically, we found that individuals with AN exhibited a lower evidence accumulation rate than HCs when presented with food-unrelated outcome-irrelevant information, while no difference was observed when presented with food-related outcome-irrelevant information. These results are consistent with a similar investigation conducted by Alfei et al. (2023) that examined hoarding patients and HCs in a PRL task. In that study, disease-related outcome-irrelevant information improved patient performance relative to a disease-unrelated condition, where their performance was inferior to that of HCs. In both cases, it appears that external information that holds personal significance activates additional cognitive resources that may not be utilized in other conditions.

Comorbidity

Individuals with eating disorders often have comorbid psychiatric conditions, including depression (up to 75%), bipolar disorder (10%), anxiety disorders, obsessive-compulsive disorder (40%), panic disorder (11%), social anxiety disorder/social phobia, post-traumatic stress disorder (prevalence varies with eating disorder), and substance abuse (15-40%) – see Woodside & Staab (2006) for further details.

In this study, we included patients with comorbidities in our sample in order to increase the generalizability of our findings to the broader psychiatric population: 16 individuals with AN were diagnosed with comorbid anxiety disorder, 8 with OCD, 1 with social phobia, and 1 with DAP; among the individuals with BN, 4 were diagnosed with mood disorder and 1 with OCD.

Importantly, our study revealed that individuals with AN who had comorbid clinical diagnoses (such as depression, anxiety, and obsessive-compulsive disorder) exhibited similar conservative reinforcement learning patterns as those without comorbidities, providing additional evidence for the specificity of the context-dependent conservative learning hypothesis for AN.

Below, we present the results obtained by fitting model M7 to the patient data, wherein the patients were stratified into two groups based on the presence or absence of comorbidities.

```
m = hddm.HDDMrl(
    data,
    depends_on={
        "a": ["diag_cat", "comorbidity", "stim"],
        "v": ["diag_cat", "comorbidity", "stim"],
        "t": ["diag_cat", "comorbidity", "stim"],
        "alpha": ["diag_cat", "comorbidity", "stim"],
        "pos_alpha": ["diag_cat", "comorbidity", "stim"],
    },
    dual=True,
    p_outlier=0.05,
    informative=True
)
```

In each instance, the 95% credibility interval encompassed zero, thus indicating a lack of credible evidence for any differences in the RLDDM parameters between patients with and without comorbid diagnoses.

Parameter a

For anorexic patients, the difference in the posterior estimates of the a parameter for the two groups, $a_{\text{comorbidity present}} - a_{\text{comorbidity absent}}$, was 0.050, 95% CI [-0.124, 0.220] in the "food" condition; the difference in the posterior estimates of the a parameter for the two groups, $a_{\text{comorbidity present}} - a_{\text{comorbidity absent}}$, was 0.097, 95% CI [-0.072, 0.256] in the "neutral" condition.

For bulimic patients, the difference in the posterior estimates of the a parameter for the two groups, $a_{\text{comorbidity present}} - a_{\text{comorbidity absent}}$, was 0.116, 95% CI [-0.195, 0.396] in the "food" condition; the difference in the posterior estimates of the a parameter for the two groups, $a_{\text{comorbidity present}} - a_{\text{comorbidity absent}}$, was 0.065, 95% CI [-0.258, 0.382] in the "neutral" condition.

Parameter α^-

For anorexic patients, the difference in the posterior estimates of the α^- parameter for the two groups, $\alpha^-_{\text{comorbidity present}} - \alpha^-_{\text{comorbidity absent}}$, was -2.071, 95% CI [-6.738, 2.446] in the "food" condition; the difference in the posterior estimates of the a parameter for the two groups, $\alpha^-_{\text{comorbidity present}} - \alpha^-_{\text{comorbidity absent}}$, was 0.195, 95% CI [-4.514, 4.900] in the "neutral" condition.

For bulimic patients, the difference in the posterior estimates of the α^- parameter for the two groups, $\alpha^-_{\text{comorbidity present}} - \alpha^-_{\text{comorbidity absent}}$, was 1.357, 95% CI [-5.061, 7.299] in the "food" condition; the difference in the posterior estimates of the a parameter for the two groups, $\alpha^-_{\text{comorbidity present}} - \alpha^-_{\text{comorbidity absent}}$, was 1.457, 95% CI [-4.393, 7.366] in the "neutral" condition.

Parameter α^+

For anorexic patients, the difference in the posterior estimates of the α^+ parameter for the two groups, $\alpha^+_{\text{comorbidity present}} - \alpha^+_{\text{comorbidity absent}}$, was 0.429, 95% CI [-3.261, 4.326] in the "food" condition; the difference in the posterior estimates of the a parameter for the two groups, $\alpha^+_{\text{comorbidity present}} - \alpha^+_{\text{comorbidity absent}}$, was 0.823, 95% CI [-3.273, 5.345] in the "neutral" condition.

For bulimic patients, the difference in the posterior estimates of the α^+ parameter for the two groups, $\alpha^+_{\text{comorbidity present}} - \alpha^+_{\text{comorbidity absent}}$, was 1.674, 95% CI [-3.723, 7.396] in the "food" condition; the difference in the posterior estimates of the a parameter for the two groups, $\alpha^+_{\text{comorbidity present}} - \alpha^+_{\text{comorbidity absent}}$, was 1.080, 95% CI [-4.487, 6.337] in the "neutral" condition.

Biased choices

In order to determine whether the conservative learning behavior observed in patients with AN during the RL task could be solely attributed to a preference for non-food-related images, irrespective of the past action-outcome history, we conducted an analysis focusing on the frequency of choices related to food in the PRL blocks. Specifically, we examined the PRL blocks in which a food image was paired with a neutral image. This analysis allowed us to explore the influence of the presence of food-related stimuli on the decision-making behavior of participants with AN.

The following table displays the proportion of times the food image was chosen when it was paired with a food-unrelated image, separately for each group.

```
## group_by: one grouping variable (di)
```

```
## summarise: now 4 rows and 2 columns, ungrouped
```

di	m
HC	0.480
AN	0.447
BN	0.493
RI	0.474

The individual proportions of food choices for each participant were analyzed using a robust Bayesian regression model (t-Student) in which the HC group was coded as the baseline group. A Beta regression model yielded similar results.

```
priors <- c(
  set_prior("student_t(4, 0, 2.5)", class = "b")
)

bmod_02 <- brm(
  bf(y ~ di, sigma ~ di),
  data = bysubj_freq,
  family = student(),
  control = list(adapt_delta = 0.99),
  prior = priors,
  backend = "cmdstan",
  warmup = 1000,
  iter = 5000,
  cores = parallel::detectCores(),
  seed = "12345",
```

```
  chains = 4,
  refresh = 0,
  silent = TRUE
)
```

Results

The results revealed a bias against food-related images, with the proportion of choices for the food-related image being 0.484, 95% CI [0.477, 0.492]. However, no group-specific bias was detected, as evidenced by the following three comparisons: AN - HC: prop = -0.002, 95% CI [-0.029, 0.026]; BN - HC: prop = 0.015, 95% CI [-0.029, 0.056]; BN - AN: prop = 0.017, 95% CI [-0.035, 0.064]; RI- HC: prop = -0.007, 95% CI [-0.031, 0.016].

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

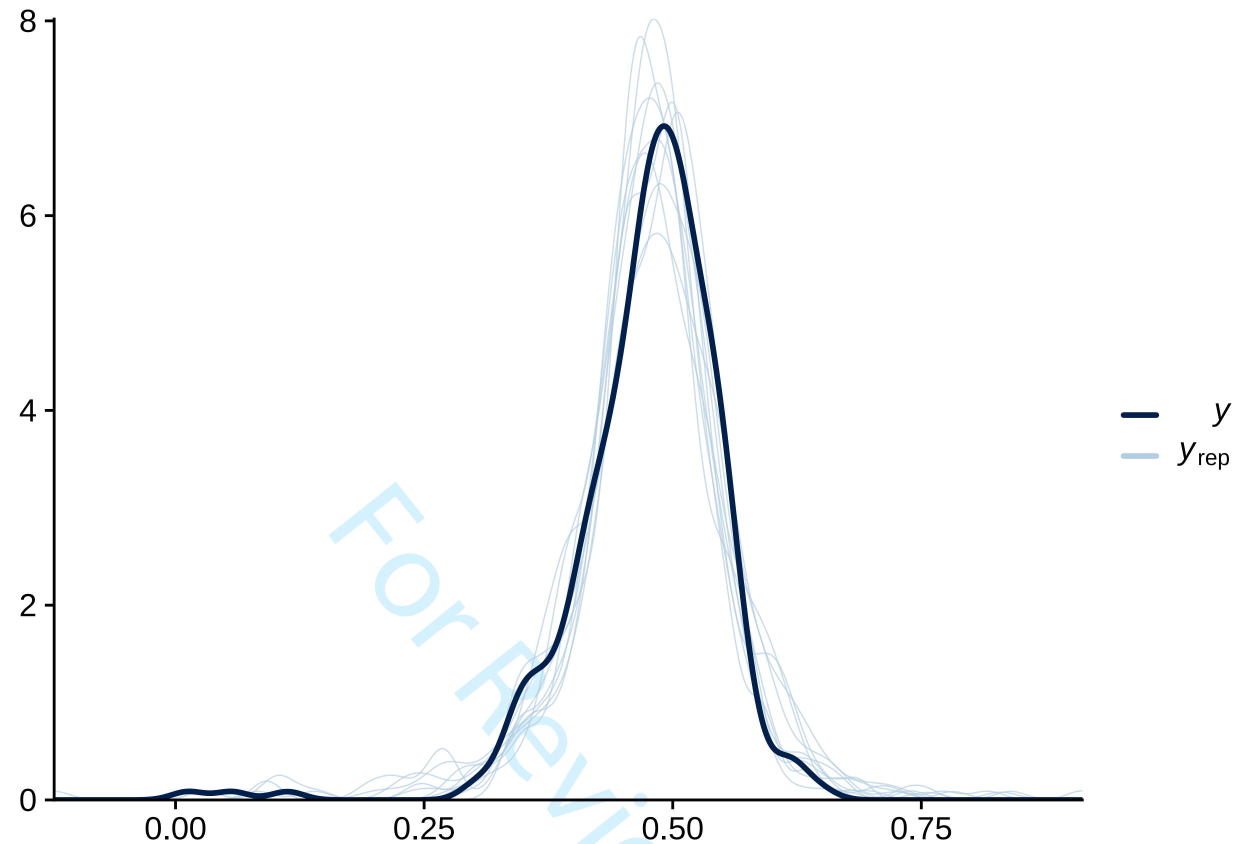
term	estimate	std.error	conf.low	conf.high
(Intercept)	0.483	0.004	0.474	0.491
sigma_(Intercept)	-2.970	0.082	-3.128	-2.807
diAN	-0.008	0.017	-0.046	0.023
diBN	0.013	0.023	-0.031	0.060
diRI	-0.007	0.012	-0.032	0.016
sigma_diAN	0.468	0.221	0.027	0.883
sigma_diBN	0.243	0.280	-0.269	0.815
sigma_diRI	0.216	0.148	-0.074	0.507

Contrasts were obtained with the `emmeans` function of the `emmeans` package:

contrast	estimate	lower.HPD	upper.HPD
AN - HC	-0.007	-0.046	0.023
BN - HC	0.014	-0.031	0.060
BN - AN	0.021	-0.032	0.079
RI - HC	-0.007	-0.032	0.016
RI - AN	0.000	-0.037	0.044
RI - BN	-0.021	-0.069	0.031

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```



Interpretation

The 95% credibility interval for the HC group was found to be non-inclusive of 0. This observation suggests that participants in the HC group exhibited a decreased likelihood of selecting the food image relative to the neutral image. In contrast, the results of the between-group comparisons failed to reveal any noteworthy differences in preferential choices between the other groups and the HC group. Moreover, no noteworthy differences were observed between the AN and BN groups.

Outcome-irrelevant learning: spatial-motor associations

Shahar et al. (2019) investigated the impact of spatial-motor associations on participants' RL. Optimal decision-making prioritizes rewards regardless of spatial-motor associations, such as the choice of response key in the previous trial. Shahar et al. (2019) found that rewards had a greater impact on the probability of choosing between two images presented in each trial when the chosen image was linked to the same response key in both the n-1 and n trials.

In order to investigate whether the likelihood of selecting 'stay' was greater for 'same' versus 'flipped' response/key mapping in our data, when contrasting rewarded and unrewarded responses, we reproduced the statistical analyses conducted by Shahar et al. (2019) and Ben-Artzi et al. (2022).

Model 1

First, we employed a 3-way interaction model that permitted the modulation of the interaction between 'same/flipped' response/key mapping × previously rewarded/unrewarded responses to vary across different groups.

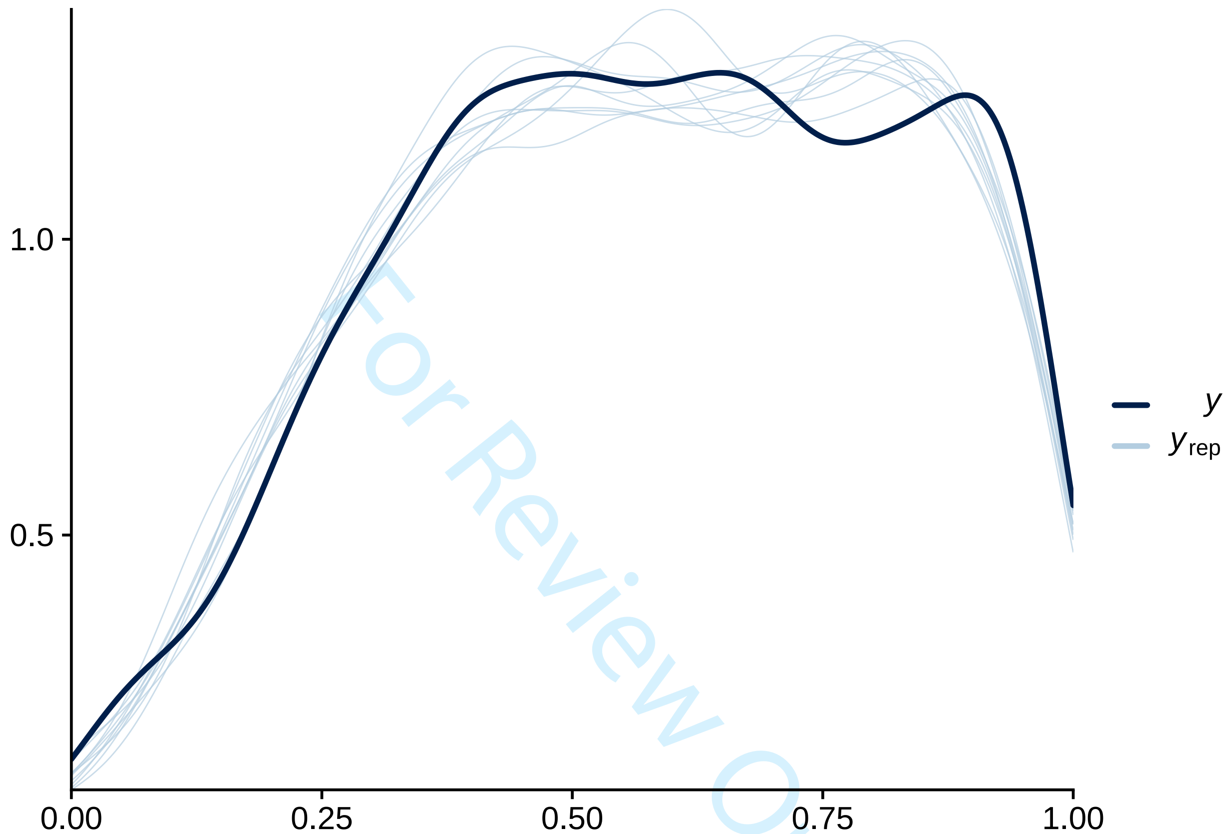
Three-way interaction

```
priors_0 <- c(
  set_prior("student_t(3, 0, 0.2)", class = "b", coef = "Intercept"),
  set_prior("student_t(3, 0, 0.2)", class = "b"),
  set_prior("student_t(3, 0, 0.2)", class = "sd"),
  set_prior("lkj(1)", class = "cor"),
  set_prior("gamma(0.01, 0.01)", class = "phi"),
  set_prior("beta(2, 2)", class = "coi"),
  set_prior("beta(2, 2)", class = "zoi")
)

mod_0 <- brm(
  stay ~ 0 + Intercept + mapping * feedback * diagnosis +
    (1 + mapping * feedback | subj_code),
  family = zero_one_inflated_beta(),
  backend = "cmdstanr",
  data = bysubj_ed,
  prior = priors_0,
  iter = 2000,
  refresh = 0,
  silent = TRUE
)
mod_0 <- add_criterion(mod_0, "loo")
```

Posterior predictive check

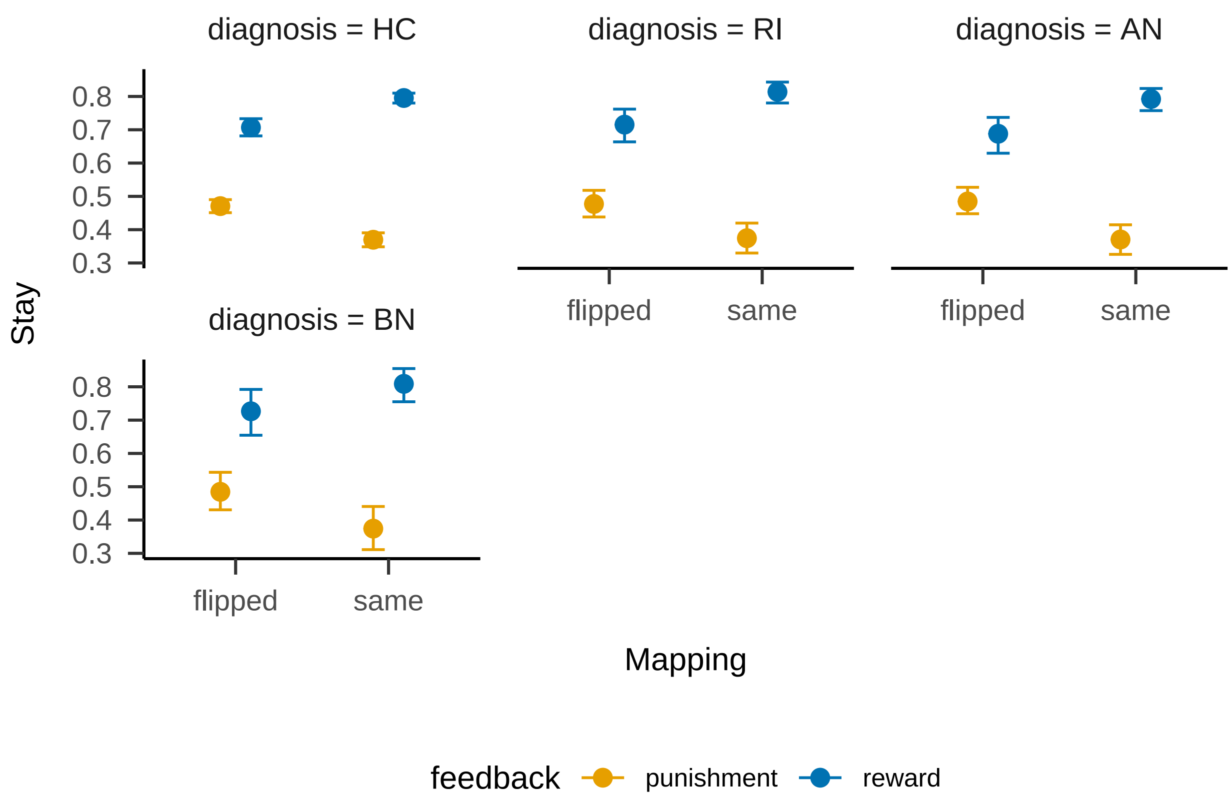
```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```



Model's coefficients

The estimations obtained for the model are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter.

term	estimate	std.error	conf.low	conf.high
(Intercept)	-0.126	0.040	-0.209	-0.053
mappingsame	-0.422	0.054	-0.530	-0.318
feedbackreward	1.011	0.082	0.866	1.177
diagnosisRI	0.025	0.084	-0.139	0.189
diagnosisAN	0.059	0.084	-0.095	0.231
diagnosisBN	0.059	0.116	-0.158	0.297
mappingsame:feedbackreward	0.905	0.066	0.782	1.040
mappingsame:diagnosisRI	-0.007	0.114	-0.238	0.211
mappingsame:diagnosisAN	-0.057	0.114	-0.280	0.173
mappingsame:diagnosisBN	-0.038	0.147	-0.328	0.256
feedbackreward:diagnosisRI	0.015	0.145	-0.259	0.304
feedbackreward:diagnosisAN	-0.155	0.157	-0.461	0.150
feedbackreward:diagnosisBN	0.042	0.189	-0.312	0.437
mappingsame:feedbackreward:diagnosisRI	0.084	0.127	-0.163	0.340
mappingsame:feedbackreward:diagnosisAN	0.134	0.132	-0.109	0.408
mappingsame:feedbackreward:diagnosisBN	0.026	0.165	-0.316	0.355



Interpretation

Based on the absence of compelling evidence for a three-way interaction, we made the decision to simplify the model by disregarding variations across groups.

Model 2

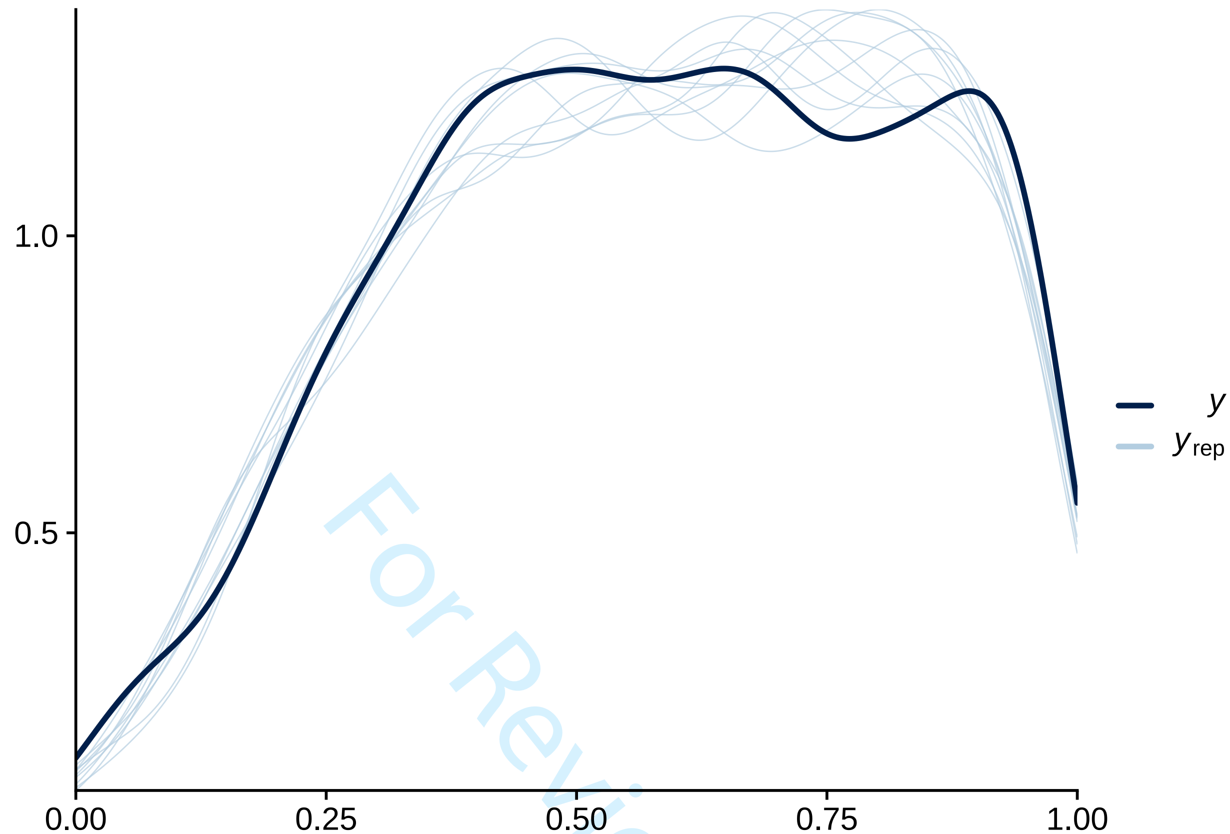
```
bcpriors <- get_prior(
  stay ~ 0 + Intercept + mapping * feedback +
    (1 + mapping * feedback | subj_code),
  family = zero_one_inflated_beta(),
  data = bysubj_ed
)

priors_1 <- c(
  set_prior("student_t(3, 0, 0.2)", class = "b", coef = "Intercept"),
  set_prior("student_t(3, 0, 0.2)", class = "b"),
  set_prior("student_t(3, 0, 0.2)", class = "sd"),
  set_prior("lkj(1)", class = "cor"),
  set_prior("gamma(0.01, 0.01)", class = "phi"),
  set_prior("beta(2, 2)", class = "coi"),
  set_prior("beta(2, 2)", class = "zoi")
)

mod_1 <- brm(
  stay ~ 0 + Intercept + mapping * feedback +
    (1 + mapping * feedback | subj_code),
  family = zero_one_inflated_beta(),
  backend = "cmdstanr",
  data = bysubj_ed,
  prior = priors_1,
  iter = 2000,
  refresh = 0,
  silent = TRUE,
  save_pars = save_pars(all = TRUE)
)
mod_1 <- add_criterion(mod_1, "loo")
```

Posterior predictive check

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```



Model's coefficients

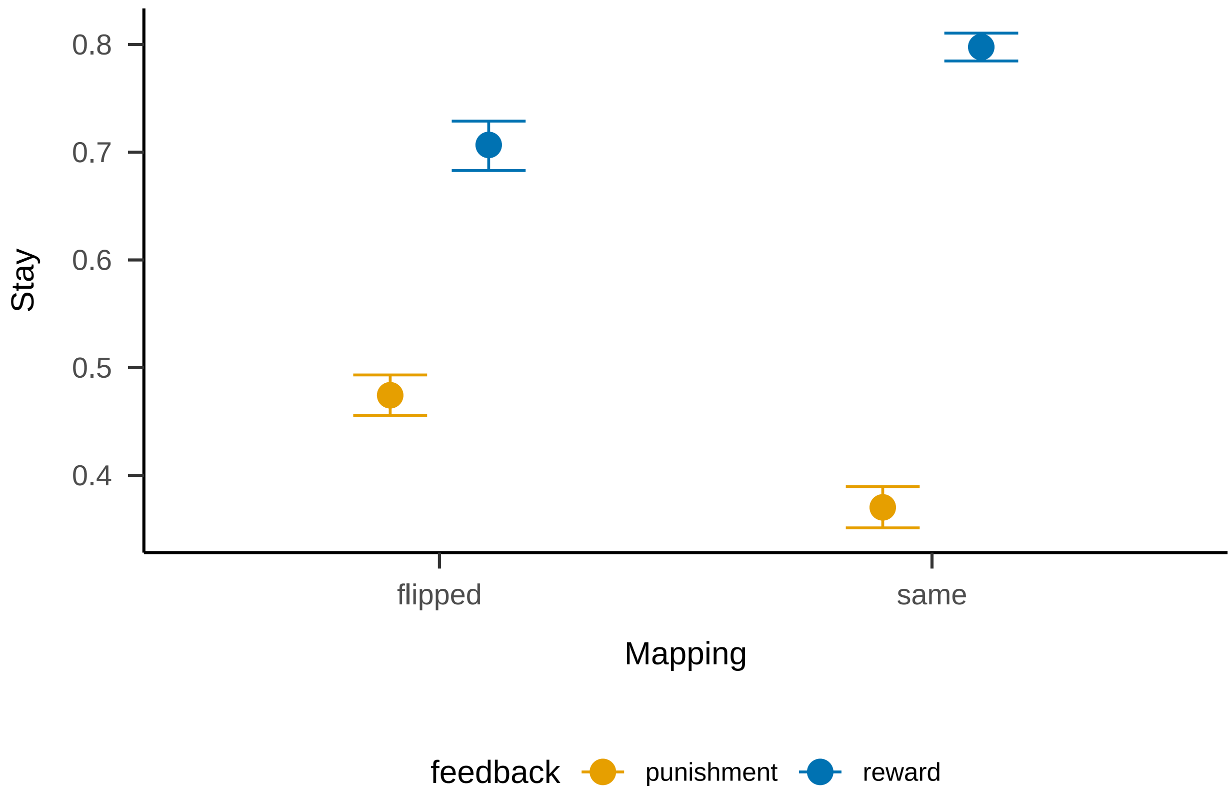
The estimated model parameters are summarized in Table below, which includes the mean, the standard error, and the lower and upper bounds of the 95% credible interval of the posterior distribution for each parameter in the posterior distribution.

term	estimate	std.error	conf.low	conf.high
(Intercept)	-0.112	0.039	-0.185	-0.033
mappingsame	-0.434	0.052	-0.532	-0.332
feedbackreward	0.993	0.076	0.843	1.144
mappingsame:feedbackreward	0.934	0.062	0.819	1.060

Interpretation

We successfully replicated the findings of Shahar et al. (2019) and Ben-Artzi et al. (2022) in the present study. Our results show robust evidence for spatial-motor outcome-irrelevant learning: the probability of choosing "stay" was higher for "same" (.427) compared to "flipped" (.218) response/key mapping when comparing previously rewarded versus unrewarded responses (posterior $\beta = 0.93$, SE = 0.06, [“HDI”]_.95 = [0.81, 1.06]; probability of direction (pd) 1.0. There was no group (HC, AN, BN, RI) × previous outcome × mapping interaction.

The figure presented below displays the predicted marginal effects of Model 2, which replicate the findings reported in Fig. 2C of Shahar et al. (2019).



Comparison with previous studies

Other recent studies have examined the issue of the domain-specificity of maladaptive associative learning in eating disorders. One task that has been specifically devised for this purpose is the two-step Markov decision task, which differentiates between automatic or habitual (model-free) and controlled or goal-directed (model-based) learning. For example, Foerde et al. (2021) and Onysk and Onysk & Seriès (2022) both conducted similar experiments using this task, with Foerde et al. (2021) comparing a monetary two-step task and a food two-step task, and Onysk & Seriès (2022) using stimuli unrelated to food or body images (pirate ships and treasure chests) with rewards associated with body image dissatisfaction. The results of these experiments showed that individuals with AN displayed a stronger preference for habitual control over goal-directed control across domains compared to healthy controls, but there were no differences in the learning rate. However, the primary aim of the two-step task is to determine whether the participants' decision-making strategy is influenced by the context or is solely based on the previous feedback received, regardless of the context. The results Foerde et al. (2021) and Onysk & Seriès (2022) show that AN patients have difficulty adapting to changing contexts compared to HCs. However, these experiments did not show any effect of the kind of context (food-related vs. food-unrelated) on decision-making in AN. More importantly, the two-step task did not reveal any difference in the learning rate of AN patients compared to HCs, as a function of the context. In contrast, by using a different method (Kool et al., 2016), our results indicate that the learning process itself, particularly the rate at which values are updated, is influenced by food-related information, even when such information is not relevant to the outcome.

Considerations on Cognitive Remediation Therapy

Cognitive Remediation Therapy (CRT) has been proposed as an adjuvant treatment targeting specific cognitive processes in several mental disorders, including eating disorders (EDs). CRT involves cognitive exercises and behavioral interventions that aim to increase central coherence abilities, reduce cognitive and behavioral inflexibility, and enhance thinking styles comprehension (Tchanturia et al., 2010). A key aspect of CRT is to avoid discussing symptoms-related themes, using neutral stimuli in cognitive and behavioral exercises to decrease drop-out rates and develop therapeutic alliance, especially with AN patients. Although some evidence suggests that CRT can be effective in improving cognitive functioning and reducing disordered eating behaviors (Tchanturia et al., 2017), a recent meta-analysis indicates that CRT was not associated with improvement in central coherence abilities, cognitive flexibility, or ED-related symptoms or comorbid psychological conditions (Hagan et al., 2020).

Trapp et al. (2022) have proposed several improvements to address practical issues encountered in cognitive remediation therapy (CRT) application. One of the proposed changes is to question the use of neutral stimuli in cognitive and behavioral exercises. The motivation for this proposal stems from Beck's cognitive theory of depression, which suggests that emotional-related factors, including cognitive biases, can impact various cognitive domains and play a pivotal role in the development and maintenance of psychopathology (Beck et al., 1987). Support to the proposal of Trapp et al. (2022) comes from evidence showing that patients diagnosed with major depressive disorder exhibit greater sensitivity to tasks involving emotional stimuli or symptoms-related stimuli, compared to healthy controls (e.g., Roiser & Sahakian, 2013). This observation aligns with the present study's finding that emotional-related factors play an essential role in cognitive processes. Therefore, we speculate that interventions designed to enhance cognitive flexibility in eating disorders and other psychological conditions should consider the influence of symptoms-related information. This can be achieved by developing cognitive and behavioral training programs that reduce the disruptive impact of cognitive biases caused by specific information. Trapp et al. (2022) propose that modifying CRT to incorporate these considerations could potentially reduce clinical symptoms more effectively. However, further investigation is necessary to confirm this hypothesis.

References

Alfei, V., Colpizzi, I., Sica, C., & Caudek, C. (2023). Domain-specificity in associative learning among hoarding patients. *Manuscript in Preparation*.

Beck, A. T., Brown, G., Steer, R. A., Eidelson, J. I., & Riskind, J. H. (1987). Differentiating anxiety and depression: A test of the cognitive content-specificity hypothesis. *Journal of Abnormal Psychology*, *96*(3), 179–183.

Ben-Artzi, I., Luria, R., & Shahar, N. (2022). Working memory capacity estimates moderate value learning for outcome-irrelevant features. *Scientific Reports*, *12*(1), 1–10.

Bernardoni, F., King, J. A., Geisler, D., Ritschel, F., Schwoebel, S., Reiter, A. M., Endrass, T., Rössner, V., Smolka, M. N., & Ehrlich, S. (2021). More by stick than by carrot: A reinforcement learning style rooted in the medial frontal cortex in anorexia nervosa. *Journal of Abnormal Psychology*, *130*(7), 736–747.

Bottesi, G., Ghisi, M., Altoè, G., Conforti, E., Melli, G., & Sica, C. (2015). The Italian version of the Depression Anxiety Stress Scales-21: Factor structure and psychometric properties on community and clinical samples. *Comprehensive Psychiatry*, *60*, 170–181.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairbum, C. G. (1987). The development and validation of the body shape questionnaire. *International Journal of Eating Disorders*, *6*(4), 485–494.

Dahlenburg, S. C., Gleaves, D. H., & Hutchinson, A. D. (2019). Anorexia nervosa and perfectionism: A meta-analysis. *International Journal of Eating Disorders*, *52*(3), 219–229.

Dotti, A., & Lazzari, R. (1998). Validation and reliability of the italian EAT-26. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *3*(4), 188–194.

Dowson, J., & Henderson, L. (2001). The validity of a short version of the Body Shape Questionnaire. *Psychiatry Research*, *102*(3), 263–271.

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. *Elife*, *10*, e65074.

Foerde, K., Daw, N. D., Rufin, T., Walsh, B. T., Shohamy, D., & Steinglass, J. E. (2021). Deficient goal-directed control in a population characterized by extreme goal pursuit. *Journal of Cognitive Neuroscience*, *33*(3), 463–481.

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, *14*(5), 449–468.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*(4), 871–878.

Geisler, D., Ritschel, F., King, J. A., Bernardoni, F., Seidel, M., Boehm, I., Runge, F., Goschke, T., Roessner, V., Smolka, M. N., et al. (2017). Increased anterior cingulate cortex response precedes behavioural adaptation in anorexia nervosa. *Scientific Reports*, *7*(1), 1–10.

Grabhorn, R., Stenner, H., Stangier, U., & Kaufhold, J. (2006). Social anxiety in anorexia and bulimia nervosa: The mediating role of shame. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *13*(1), 12–19.

Guarda, A. S., Schreyer, C. C., Boersma, G. J., Tamashiro, K. L., & Moran, T. H. (2015). Anorexia nervosa as a motivated behavior: Relevance of anxiety, stress, fear and learning. *Physiology & Behavior*, *152*, 466–472.

Hagan, K. E., Christensen, K. A., & Forbush, K. T. (2020). A preliminary systematic review and meta-analysis of randomized-controlled trials of cognitive remediation therapy for anorexia nervosa. *Eating Behaviors*, *37*, 101391.

Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., & Liebowitz, M. R. (1992). Assessment of anxiety in social interaction and being observed by others: The social interaction anxiety scale and the social phobia scale. *Behavior Therapy*, *23*(1), 53–73.

Kaestner, D., Loewe, B., & Gumz, A. (2019). The role of self-esteem in the treatment of patients with anorexia nervosa–a systematic review and meta-analysis. *International Journal of Eating Disorders*, *52*(2), 101–116.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off? *PLoS Computational Biology*, *12*(8), e1005090.

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, JAGS, and stan.*

Lang, P. J., Bradley, M. M., Cuthbert, B. N., et al. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* NIMH, Center for the Study of Emotion & Attention Gainesville, FL.

Lombardo, C. (2008). Adattamento italiano della multidimensional perfectionism scale (MPS). *Psicoterapia Cognitiva e Comportamentale*, *14*(3), 31–46.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470.

Onysk, J., & Seriès, P. (2022). The effect of body image dissatisfaction on goal-directed decision making in a population marked by negative appearance beliefs and disordered eating. *Plos One*, *17*(11), e0276750.

Pedersen, M. L., & Frank, M. J. (2020). Simultaneous hierarchical bayesian parameter estimation for reinforcement learning and drift diffusion models: A tutorial and links to neural data. *Computational Brain & Behavior*, *3*, 458–471.

Perez Santangelo, A., Ludwig, C. J., Navajas, J., Sigman, M., & Leone, M. J. (2022). Background music changes the policy of human decision-making: Evidence from experimental and drift-diffusion model-based approaches on different decision tasks. *Journal of Experimental Psychology: General*.

Pleple, A., Lalanne, C., Huas, C., Mattar, L., Hanachi, M., Flament, M., Carchon, I., Jouen, F., Berthoz, S., & Godart, N. (2021). Nutritional status and anxious and depressive symptoms in anorexia nervosa: A prospective study. *Scientific Reports*, *11*(1), 1–12.

Prezza, M., Trombaccia, F. R., & Armento, L. (1997). La scala dell'autostima di rosenberg: Traduzione e validazione italiana. *Giunti Organizzazioni Speciali*.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Raven, Jean. (2003). *Raven progressive matrices*. Springer.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for raven's progressive matrices and vocabulary scales. Section 3: The standard progressive matrices*. Oxford Psychologists Press.

Rescorla, R. A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current Research and Theory*, 64–99.

Roiser, J. P., & Sahakian, B. J. (2013). Hot and cold cognition in depression. *CNS Spectrums*, *18*(3), 139–149.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.

Shahar, N., Moran, R., Hauser, T. U., Kievit, R. A., McNamee, D., Moutoussis, M., Consortium, N., & Dolan, R. J. (2019). Credit assignment to state-independent task representations and its relationship with model-based decision making. *Proceedings of the National Academy of Sciences*, *116*(32), 15871–15876.

Sica, C., Musoni, I., Bisi, B., Lolli, V., & Sighinolfi, C. (2007). Social phobia scale e social interaction anxiety scale: Traduzione e adattamento italiano. *Bollettino Di Psicologia Applicata*, *252*, 59–71.

Swinbourne, J. M., & Touyz, S. W. (2007). The co-morbidity of eating disorders and anxiety disorders: A review. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association*, *15*(4), 253–274.

Tchanturia, K., Davies, H., Reeder, C., & Wykes, T. (2010). *Cognitive remediation therapy for anorexia nervosa*. King's College London, University of London.

Tchanturia, K., Giombini, L., Leppanen, J., & Kinnaird, E. (2017). Evidence for cognitive remediation therapy in young people with anorexia nervosa: Systematic review and meta-analysis of the literature. *European Eating Disorders Review*, *25*(4), 227–236.

Trapp, W., Heid, A., Röder, S., Wimmer, F., & Hajak, G. (2022). Cognitive remediation in psychiatric disorders: State of the evidence, future perspectives, and some bold ideas. *Brain Sciences*, *12*(6), 683.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, 14.

Woodside, B. D., & Staab, R. (2006). Management of psychiatric comorbidity in anorexia nervosa and bulimia nervosa. *CNS Drugs*, *20*, 655–663.