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

Corrado Caudek1, Ilaria Colpizzi2, Lisa Guidi3, Sara Danti3, Silvia Lucchesi3, Elisa Giusti3, Martina Di Meglio3, Donatella Ballardini4, Chiara Mazzoni4, Caterina Pieraccioli5, Romana Schumann4, & Claudio Sica2

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](mailto:corrado.caudek@unifi.it)

*Word count:* 8746

**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 ( 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) 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.

We compared the characteristics of the clinical sample with the controls by administering the following scales: the Eating Attitude Test-26 (EAT-26; Garner et al., 1982), 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 can be found in the Supplementary Information (SI).

**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 mentioned earlier. In a subsequent session, participants completed the PRL task and were subsequently provided with a debriefing (for details of the study's procedure and methods, please refer to the 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, please 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.

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).

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, please 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.

**Estimating outcome-irrelevant learning**

**Reinforcement learning and drift diffusion modeling**. 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 and parameters are based on the diagnostic group. In Model M4, the and 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 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 parameter. Model M7 builds upon M6 by considering that the 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 , , , , and are conditioned on both diagnostic groups and the presence of food-related or food-unrelated outcome-irrelevant information.

The winning model was subsequently better estimated using 13000 traces and a 3000 burn-in (Kruschke, 2014). 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 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 () 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 (, ) 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 () 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 () for food-related and food-unrelated information () 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 () and negative () 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 () between the two types of information in individuals with AN. However, we did not find any credible differences in the learning rates for punishments () 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, please 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, please 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 ( 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 ( and 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 ( 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.

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., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

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.

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.

Fitzsimmons-Craft, E. E., Accurso, E. C., Ciao, A. C., Crosby, R. D., Cao, L., Pisetsky, E. M., et al.others. (2015). Restrictive eating in anorexia nervosa: Examining maintenance and consequences in the natural environment. *International Journal of Eating Disorders*, *48*(7), 923–931.

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., & 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.

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.

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.

King, J. A., Korb, F. M., Vettermann, R., Ritschel, F., Egner, T., & Ehrlich, S. (2019). Cognitive overcontrol as a trait marker in anorexia nervosa? Aberrant task-and response-set switching in remitted patients. *Journal of Abnormal Psychology*, *128*(8), 806.

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. (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.

Matera, C., Nerini, A., & Stefanile, C. (2013). The role of peer influence on girls’ body dissatisfaction and dieting. *European Review of Applied Psychology*, *63*(2), 67-74.

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.

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.

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.

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

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.

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.

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.

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

Stöber, J. (1998). The frost multidimensional perfectionism scale revisited: More perfect with four (instead of six) dimensions. *Personality and Individual Differences*, *24*(4), 481–491.

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.

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

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 () of the DDMRL across the four participants’ groups.

Figure 1.  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 () of the DDMRL, for food-related (top row) and food-unrelated (bottom row) outcome-irrelevant information.

Figure 2.  Plots of the posterior distribution of the group effect for parameter \alpha^+ (\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 () of the DDMRL, for food-related (top row) and food-unrelated (bottom row) outcome-irrelevant information.

Figure 3.  Plots of the posterior distribution of the group effect for parameter \alpha^- (\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 () for parameter 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 (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 () of the DDMRL across the four participants’ groups.

Figure 5.  Plots of the posterior distribution of the domain-specificity effect for parameter \alpha^+ (\alpha^+_{\text{food-related}} - \alpha^+_{\text{food-unrelated}}) of the DDMRL across the four participants’ groups.