UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
SCIENZE CLINICHE

CICLO XXXVI

COORDINATORE Prof. Lorenzo Cosmi

Cognitive Inflexibility as Transdiagnostic Risk Factor in Psychopathological Disorders

Settore Scientifico Disciplinare M-PSI/08

Dottoranda
Dott.ssa Ilaria Colpizzi

Tutore
Prof. Claudio Sica

Coordinatore
Prof. Lorenzo Cosmi

Prof. COSMI LORENZO
Matr. 999314

Anni 2020/2023

Abstract

Cognitive inflexibility (CI), an inability to adapt to environmental changes, is posited as a *transdiagnostic* factor maintaining symptoms in multiple psychopathologies. Prior focus on merely symptom-based classifications may inhibit the comprehension and treatment of mental disorders, potentially missing their underlying causes. The Research Domain Criteria (RDoC) offers a holistic approach, emphasizing *transdiagnostic* factors and integrating insights from various disciplines.

CI, aligning with the *Cognitive Control* construct in the RDoC, has been associated with Eating Disorders, especially Restrictive-type Anorexia Nervosa (AN-R), though with inconclusive evidence. Traditional diagnostic methods for eating disorders face challenges, highlighting the need for a *transdiagnostic* perspective. The current research aimed to explore this direction by elucidating the role of CI in AN-R and, to a lesser extent, in Bulimia Nervosa (BN). Addressing critical gaps is essential to achieve this objective, particularly: The nature of CI - whether the deficit is domain-general or domain-specific (**Study 1**), the impact on cognitive processes due to variations in symptom intensity among AN-R patients, as well as the potential deficit of cognitive flexibility in BN (**Study 2**), and CI multifaceted characterization - specifically, whether the compromised ability is reversal learning, set/task-switching, or both (**Study 3**).

Findings from the current study, using the Probabilistic Reversal Learning task, revealed a cognitive flexibility deficit in AN-R and BN compared with controls. This deficit was domain-specific for patients with low to medium symptom intensity and domain-general for those with high intensity. Reversal learning, compared to set/task switching (measured using the Task Switching and Wisconsin Card Sorting Test), was identified as the impaired ability. Data were analyzed using Computational Modeling, offering robust techniques. These findings suggest integrating cognitive flexibility interventions with standard therapies like Cognitive-Behavioral Therapy for AN-R and BN.

In conclusion, CI emerges as a potential consistent maintaining factor in AN-R, and potentially in BN. The results obtained pave the way for future investigations into other disorders where the role of CI is still debated.

Contents

List of Figures

List of Tables

Premise

The complex relationship between cognitive mechanisms and symptoms of psychopathology has consistently intrigued the scientific community. Recently, Cognitive Flexibility (CF), defined as the capacity to adapt to changing environmental demands, has gained increased attention. Morris & Mansell (2018) proposed a CF deficit as a potential *transdiagnostic* factor, highlighting its possible influence in intensifying and sustaining symptoms across a multitude of psychological disorders.

Diverging from a traditional symptom-based classification, an emphasis on *transdiagnostic* elements—cognitive and behavioral attributes, among other factors, observed across a range of disorders—might foster a deeper understanding and subsequently refined therapeutic interventions. Notably, renowned diagnostic systems, specifically the DSM-5-TR (American Psychiatric Association, 2022) and the ICD-11 (World Health Organization, 2018), while monumental in their own right, have shown potential limitations in their predictive and therapeutic scopes, possibly attributable to gaps in comprehending the core etiologies of mental disorders (Watkins, 2015).

The Research Domain Criteria (RDoC) initiative—a paradigm shift promoting a holistic and integrated perspective on mental disorders-was recently proposed. The RDoC framework integrates knowledge from various scientific fields such as biology, psychology, neuroscience, and genetics. The emphasis here is on uncovering basic mechanisms—spanning cognitive, emotional, and behavioral dimensions—that traverse diagnostic boundaries. Such an integrated perspective offers a detailed insight into psychopathology, potentially paving the way for tailored therapeutic treatments (Watkins, 2015).

The RDoC classification system places CI within the ambit of the *Cognitive Control*, a construct located under the broader *Cognitive* domain. Thus, this research undertakes an exhaustive investigation into the *transdiagnostic* importance of CI within the field of psychopathology. The initial focus is on the field of Restrictive-type Anorexia Nervosa (AN-R) and, to a lesser extent, Bulimia Nervosa (BN). The traditional diagnostic frameworks, which primarily rely on symptomatology for classifying eating disorders, often straggle with challenges such as ambiguous diagnostic overlaps and indistinct diagnostic boundaries (Garcia-Burgos, 2022).

Given these considerations, delving deeper into the *transdiagnostic* characteristics of eating disorders is essential. While the association of CI with eating disorders, particularly AN-R, has been a significant area of academic research, the current literature presents a variety of differing conclusions.

The lack of consensus may be attributed to several limitations: (1) CI has traditionally been examined as a generalized deficit (domain-general). However, it is possible that the deficit manifests primarily in response to symptom-related cues (domain-specific) (Hitchcock, Fried, & Frank, 2022); (2) Often, heterogeneity in symptom severity in AN is overlooked, even though different severity levels might be associated with unique cognitive patterns (Davis, Walsh, Schebendach, Glasofer, & Steinglass, 2020); (3) CI is frequently regarded as a singular entity, neglecting its multifaceted nature, which encompasses both *reversal learning* and *set/task switching* abilities (Wildes, Forbes, & Marcus, 2014).

These theoretical limitations naturally lead to methodological issues. Foremost among these is the variability in measurement approaches and the subsequent analysis methods. The challenge of identifying the optimal behavioral markers for evaluating performance persists.

The study aims to shed light on the role of CI in AN-R and BN. Results suggest a discernible deficit in cognitive flexibility in AN-R patients. By using the Probabilistic Reversal Learning task, it was observed that the manifestation of the CI deficit varies based on symptom severity. Specifically, for patients with milder symptoms, the CI deficit was found to be domain-specific. In contrast, for patients with severe symptoms, the deficit was domain-general. Moreover, the *reversal learning* aspect of CI was found to be impaired in this particular patient population.

Deficits in cognitive flexibility were also found in BN, supporting the *transdiagnostic* nature of CI.

Finally, the adoption of advanced computational models for data interpretation yielded profound insights, emphasizing the value of these models in analyzing reinforcement learning tasks (Haynos, Widge, Anderson, & Redish, 2022).

These results could carry significant clinical implications. Deficits in cognitive flexibility may hinder the success of treatments, such as Cognitive-Behavioral Therapy for AN-R and BN (Fairburn, Cooper, & Shafran, 2003). Thus, a two-fold strategy that combines targeted interventions for this deficit with traditional psychotherapeutic methods seems highly advisable (Hagan, Christensen, & Forbush, 2020; Tchanturia, Lounes, & Holttum, 2014). For these interventions to be effective, it is essential to conduct thorough assessments of CI using specialized tools that recognize its complex nature.

In summary, the study highlights the potential significance of CI in understanding and potentially addressing AN-R and BN symptoms. While these findings provide clarity, they necessitate further validation from additional research.

Given these insights, it could be interesting to extend the study of CI to other psychopathologies. Especially those disorders where CI has been investigated but the evidence remain inconclusive. This could provide further evidence supporting the *transdiagnostic* nature of CI.

Chapter 1

General Introduction

A comprehensive understanding of psychopathology requires an in-depth examination of its causes and the underlying mechanisms. Recently, there has been a heightened focus on investigating *transdiagnostic* processes, which manifest across a range of disorders. Cognitive flexibility, the ability to adapt thoughts and behaviors to environmental demands, is central to this discussion. Its absence is linked to disorders like schizophrenia and obsessive-compulsive disorder. However, the role of cognitive flexibility in other disorders, especially eating disorders like Anorexia Nervosa, remains inconsistent and needs further research.

1.1 Understanding Transdiagnostic Factors in Psychopathology

A deeper understanding of psychopathology requires a detailed evaluation of its underlying causes and the maintaining mechanisms (Morris & Mansell, 2018). Recent studies, covering fields from genetics to psychology, have questioned the current diagnostic system. This system, which is mainly based on assessing the presence or absence of specific symptoms or patterns, often overlooks the underlying factors that may cause or maintain psychopathology (Visu-Petra & Mărcuș, 2019). These risk and maintaining factors are known as *transdiagnostic* processes. Shifting the focus on *transdiagnostic* factors, aligns with the principles of the Research Domain Criteria project (RDoC, Kozak & Cuthbert, 2016), an initiative led by the U.S. National Institute of Mental Health (NIMH). Instead of traditional symptom-based classifications like those in Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition-Text Revised (DSM-5-TR, American Psychiatric Association, 2022), the RDoC emphasizes dimensions grounded in cognitive, neurophysiological, and behavioral research, identifying them as *transdiagnostic* factors. This shift represents a substantial change in how we understand psychopathology and clinical psychology. While this approach holds promise, further research is underway to delve deeper into these *transdiagnostic* factors. The ultimate goal of the RDoC is to transition from the categorical/dimensional perspective found in the DSM-5-TR to a model rooted in *transdiagnostic* factors, which can guide therapeutic interventions.

What does *transdiagnostic* truly mean?

This term broadly applies to a diverse array of factors, extending beyond cognitive processes like attention and memory, and behavioral processes (*e.g.*, avoidance). It includes a variety of elements such as emotional, social, and environmental factors (*e.g.*, stress resilience, social interactions, lifestyle habits), all observed across a range of diagnostic categories.

In other words, it refers to mechanisms or traits that are not limited to a single diagnosis, but can be found in multiple psychopathological conditions (Harvey, Watkins, & Mansell, 2004). These mechanisms underpin symptoms observed, and thus play a role in the onset and the persistence of the disorder (Visu-Petra & Mărcuș, 2019). Hence, understanding the *transdiagnostic* factors is essential for a deeper comprehension of psychopathology functioning and for developing more effective clinical interventions.

According to this approach to psychopathology, Cognitive Inflexibility (CI) has been posited as a key *transdiagnostic* mechanism that underlies and sustains various symptom patterns (Giommi et al., 2023). CI is the inability to adapt to changing environments, resulting in rigid and unhelpful thought patterns. This rigidity can hinder positive changes, thereby perpetuating symptoms (Diamond, 2013). This impairment in adaptability underscores the significance of understanding the role of this cognitive process in psychopathology (Kashdan & Rottenberg, 2010).

1.2 Cognitive Flexibility: An Overview

Cognitive flexibility (CF) is a core component of Executive Functions that emerges later in development, building upon working memory and inhibitory control (Davidson, Amso, Anderson, & Diamond, 2006; Garon, Bryson, & Smith, 2008). CF involves the dynamic ability to shift way of thinking or perspective, whether this means changing spatial viewpoints or adapting interpersonal understandings. This shift demands the deactivation of a current perspective and the concurrent activation of an alternative one. Beyond mere perspective shifting, CF embodies the broader capacity for innovative thinking, adaptability to changing circumstances, the ability to recognize and correct errors, and the tendency to exploit unexpected opportunities. It aligns closely with creative thought processes and cognitive operations like task switching, set shifting, and the aptitude for reversal learning – adapting to changes in stimulus-response associations. In essence, cognitive flexibility is the antithesis of cognitive rigidity (Diamond, 2013).

Expanding on this concept, CF is not a singular, monolithic ability but rather a composite of various cognitive capacities. Two key capabilities under this umbrella are attention set-shifting and reversal learning. Attention set-shifting, also known as task-switching, describes the cognitive ability to shift focus between different attributes of a stimulus or transition between distinct cognitive tasks. Reversal learning, on the other hand, entails adjusting stimulus-response associations according to feedback, demonstrating the ability to both discard and acquire new associations.

Research has indicated that these abilities, while under the CF concept, are distinct both behaviorally and neurally. Set/task-switching is associated with neural activity in areas like the Ventromedial Prefrontal Cortex, Anterior Cingulate Cortex, and regions in the Posterior Temporal and Parietal areas. This skill develops with age, so tasks that demand a change in perception can be challenging for children around 5 years old (Diamond, 2005).

Reversal learning, conversely, is linked with the Orbitofrontal Cortex and Ventral Striatum. Remarkably, this ability emerges quite early, with children as young as 2.5 years successfully completing reversal tasks. This suggests that the capacity to update stimulus-response patterns develops earlier than the ability to alter cognitive perspectives on stimuli-set/task-switching (Perner & Lang, 2002).

Behaviorally, these components guide different strategies, suggesting that they might have distinct roles in facilitating adaptive behaviors.

With age, CF shows a clear pattern: it improves during childhood and declines in later years. During early adulthood, these abilities typically reach their highest level, suggesting optimal performance on CF tasks. Both older adults and children display greater variability in the speed of task-switching or strategy changes, particularly when dealing with different stimuli, compared to those in their younger adult years. However, even though older adults might switch tasks at a speed similar to children, they consistently perform these tasks accurately (Cepeda, Kramer, & Gonzalez de Sather, 2001; Diamond, 2016).

The discussion surrounding CF not only addresses whether it is a single or multifaceted construct but also delves into its intrinsic nature. This debate arises from a more comprehensive question that encompasses all cognitive processes: Do cognitive deficits originate from broad neural dysfunctions (domain-general), or are they tied to particular contexts (domain-specific), indicating situation-specific rather than generalized impairments? This differentiation was largely overlooked in CF research, leading to inconsistent results. However, this perspective has recently been investigated within the CF domain (Caudek, Sica, Cerea, Colpizzi, & Stendardi, 2021; Hitchcock et al., 2022). Resolving this debate is crucial for improving both the assessment and treatment of CF.

Indeed, following the idea of CF as a *transdiagnostic* sustaining factor, it has been suggested as a potential therapeutic target (Tchanturia, Davies, Reeder, & Wykes, 2010). Enhancing CF and reducing rigidity could facilitate the adoption of more adaptive behavioral strategies, fostering healthier behaviors and thought patterns that may alleviate symptoms (Davis et al., 2020). However, to develop effective therapeutic interventions, it is essential to understand CF nature, specifically whether it is a unitary or multifaceted construct, and whether the deficit is domain-general or domain-specific.

To deal with these theoretical challenges, it is imperative to start with methodological considerations.

Given this perspective, research instruments and methodologies must be carefully chosen and tailored (Hagan et al., 2020). To truly understand CF, researchers should employ tools specifically designed to probe each facet of CF individually. Relying on generic or overarching methodologies can obscure the distinct attributes of each component, potentially leading to oversimplified or inaccurate conclusions. Moreover, it is crucial to develop specific tasks that can discern between the domain-general and the domain-specific nature of the deficit. In short, meticulous research design is the way for clearer insights into the intricate world of CF.

1.2.1 Measuring Cognitive Flexibility

As previously noted, CF is a multifaceted concept, encompassing a range of cognitive skills. Predominantly, this includes abilities like set/task switching and reversal learning. Given the distinct nature of these capabilities, it is essential to use tasks tailored to evaluate each specific skill.

Various psychological assessments are employed to measure CF. Among these are Fluency tasks, which encompass Design, Verbal, and Category fluency. Widely recognized task-switching and set-shifting tasks include the Wisconsin Card Sorting Test (Grant & Berg, 1993), the Trial Making Test (Reitan, 1958), the Intra/Extra-dimensional set-shifting task (Robbins et al., 1998) and standard Task-switching paradigms. Others like the Dimensional Change Card Sort Test, the CatBat (Tchanturia, Campbell, Morris, & Treasure, 2005), and the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997) are also commonly used. Notably, these tasks primarily evaluate the set/task-switching aspect of CF.

These tasks predominantly assess the ability to shift attention between different stimulus features. For instance, in the Wisconsin Card Sorting Test, participants are presented with four fixed cards and a separate target card. They must discern the underlying rule that determines how the target card corresponds with one of the fixed cards, based on specific visual attributes like color, shape, or number. As the test progresses, this rule changes multiple times, causing the criteria for selection to shift between these visual attributes (Nyhus & Barceló, 2009).

The Brixton Spatial Anticipation Test evaluates the ability to detect and switch rules based on changing visual sequences. It consists of 56 images with a shifting circle position, and the participant must identify the underlying rule governing these changes (Van Den Berg et al., 2009).

The Intra/Extra-dimensional set-shifting task assesses both cognitive flexibility and attention shifts. Participants are required to learn and then shift their attention between different sets of stimuli based on feedback. Initially, the task focuses on changes within the same stimuli type (*intra-dimensional*) and later on shifting attention to entirely new stimuli categories (*extra-dimensional*). The task provides insight into the ability to adapt and change their cognitive focus in response to changing rules or environments (Jazbec et al., 2007).

The Trail Making Test, on the other hand, assesses various cognitive functions through two tasks: TMT-A, which tests rote memory by having participants connect numbered circles sequentially, and TMT-B, which adds complexity by alternating between numbers and letters, gauging executive functioning and task-switching ability (Bowie & Harvey, 2006). Notably, tasks like these can be adapted to use word-based stimuli, as seen in the CatBat Test (Tchanturia et al., 2005).

The best-known word-related tasks are the Fluency tasks. A fluency task is designed to measure the ability to produce numerous verbal responses within a specified category or based on particular criteria in a limited time frame. This allows to assess the speed and flexibility of thought, semantic memory access, and creativity (Shao, Janse, Visser, & Meyer, 2014).

In other tests, like the Task-switching paradigms for adults and the Dimensional Change Card Sort Test for children, participants must adjust their strategy based on explicit task instructions. During these tasks, participants are presented with two stimuli and need to follow a specific rule to choose correctly between them. After a set number of trials, the rule is changed, leading to a new task with a distinct guiding principle. The goal is to measure the challenges participants encounter when shifting from one task rule to another (Monsell, 2003).

Additionally, there is growing interest in evaluating reversal learning capabilities using Probabilistic Reversal Learning paradigms (Cools, Clark, Owen, & Robbins, 2002). These paradigms evaluate the ability to modify behavioral strategies in response to changing environments. In such settings, participants are presented with two stimuli, each associated with different probabilities of receiving a reward. The goal is to consistently select the stimulus with the higher reward probability. However, after a series of trials, the reward probabilities for the stimuli can unpredictably switch. Since these shifts are unpredictable, participants must adapt according to the feedback they receive (Monni, Scandola, Hélie, & Scalas, 2023).

At their core, these tasks test the ability to adapt thinking or perspective in response to changing conditions. This emphasizes the intricate relationship between inhibition, attention, and working memory, and the complex and demanding nature of CF.

Conclusively, this review underlines CF as a complex construct covering diverse cognitive skills. Given the vast array of tools employed, which may lead to inconsistent evidence, it is crucial to use specialized, tailored tasks to gain a detailed understanding of each specific component of CF.

Further insights on the assessment tools for CF can be found in **Study 3**.

1.3 Cognitive Inflexibility as Transdiagnostic Factor in Mental Disorders

The role of CI as a *transdiagnostic* factor in the onset and maintenance of various psychopathologies is well-documented. Numerous studies have highlighted its occurrence across a spectrum of conditions including Psychotic Disorders (Orellana & Slachevsky, 2013;

Waltz, 2017), Mood Disorders (Mukherjee, Filipowicz, Vo, Satterthwaite, & Kable, 2020), Substance Abuse (Hekmat, Mehrjerdi, Moradi, Ekhtiari, & Bakhshi, 2011; Verdejo-Garcia et al., 2015), Anxiety (Park & Moghaddam, 2017), Post-Traumatic Stress Disorder (Ben-Zion et al., 2018; Popescu, Popescu, DeGraba, & Hughes, 2023) and Eating Disorders (Miles, Gnatt, Phillipou, & Nedeljkovic, 2020; Tchanturia et al., 2011). Although each of these disorders has distinct diagnostic characteristics, they exhibit common abnormalities in brain structures, such as reduced gray matter in specific regions, and functional irregularities. These brain alterations are generally linked to cognitive inflexibility. Cognitive flexibility is crucial for overall individual well-being and greatly influences quality of life (Kashdan & Rottenberg, 2010). Essentially, deficits in this ability have been linked to a broad array of neurological and psychiatric disorders, underlining its crucial role in cognitive performance (Dajani & Uddin, 2015).

However, the evidence for a deficit in cognitive flexibility varies in both quantity and quality among the aforementioned disorders. It is clearly observed in conditions like schizophrenia, other psychotic spectrum disorders, and obsessive-compulsive disorder, where CI often exacerbates symptoms (Caudek, Sica, Marchetti, Colpizzi, & Stendardi, 2020; Chamberlain, Fineberg, Blackwell, Robbins, & Sahakian, 2006; Orellana & Slachevsky, 2013). Evidence are more ambiguous for mood disorders, substance use disorders, anxiety disorders, and eating disorders. This uncertainty for mood, substance use, and anxiety disorders might be due to a limited number of relevant studies. In contrast, for eating disorders, especially Anorexia Nervosa (AN), there is an extensive body of research on CI. Nonetheless, findings from these studies often diverge, leading to inconsistent evidence.

Despite numerous studies, the relationship between eating disorders and CI remains ambiguous. Exploring the reasons for this discrepancy is intriguing; understanding the basis of such divergence in eating disorders, specially AN, could also benefit the study of CI in other disorders where similar questions remain unanswered.

1.4 Cognitive Inflexibility in Eating Disorders: An Overview

The relationship between CI and Eating Disorders is comprehensively elucidated by Walsh's model (Walsh, 2013). While originally focused on AN, this model has been expanded to offer insights into the role of CI in a broader spectrum of eating disorders. The model conceptualizes maladaptive eating behaviors as originating from reinforcement learning (or operant conditioning) mechanisms. It delineates a progression where behaviors, such as dieting that initially yield rewarding outcomes, are repeated due to both positive (*e.g.*, increased self-esteem, reflecting self-control and personal achievement) and negative reinforcement (*e.g.*, alleviating negative emotions and thereby coping with emotional challenges), creating a stimulus-response association. Over time, these behaviors become automatic and habitual, persisting even in the absence of immediate rewards. This habit formation reflects a shift from behaviors that are reward-sensitive to those less sensitive to rewards and more resistant to change, becoming inflexible. In AN, behaviors such as dieting, initially driven by action-outcome learning with the goal of weight loss, gradually transform into deeply ingrained habits through stimulus-response learning. Consequently, these behaviors develop an almost compulsive nature, being carried out irrespective of direct or obvious rewards, and can lead to dangerous outcomes.

The model extends to Bulimia Nervosa (BN), encompassing both dietary restrictions and binge/purge behaviors. These behaviors in BN are theorized to follow a trajectory similar to AN, evolving from initially rewarding actions to compulsive habits. In Binge Eating Disorder (BED), despite its differences from AN, the model suggests that binge-eating episodes become a compulsive, addiction-like habit. This transition is attributed to the past rewarding nature of these behaviors (*e.g.*, emotion regulation) and the established strong stimulus-response associations (Banca, Harrison, Voon, & Brand, 2016). This pattern mirrors that observed in compulsive-based psychological disorders, including Obsessive-Compulsive and Substance Use Disorders (Gillan & Robbins, 2014).

Consequently, Walsh's model highlights the pivotal role of CI in maintaining symptoms of various eating disorders (Walsh, 2013). It shows how CI supports the shift from conscious, reward-oriented actions to automatic, habit-based behaviors. The model outlines an alteration in reinforcement learning processes, intimately linked to cognitive flexibility, characterized by the crystallization of specific behaviors, surpassing goal-directed behaviors that are typically adaptable and responsive to environmental feedback. In the context of eating disorders, these flexible behaviors are disrupted, underscoring the importance of addressing CI in therapeutic interventions.

Although this model was developed from a thorough review of previous studies, the evidence in the literature regarding a deficit in cognitive flexibility in eating disorders, particularly AN, remains inconclusive.

For instance, a study by Adoue et al. (2015) used a Probabilistic Reversal Learning (PRL) task to assess cognitive flexibility deficits in AN patients. They compared the number of perseverative errors and *switch errors* between AN patients and healthy controls, finding no performance differences. This observation was replicated by Bernardoni et al. (2018a), who, using the same metrics in a PRL task, also identified no differences between the two groups. However, upon analyzing their data with an advanced computational model, they noted that AN patients had a faster learning rate during punished trials, which enhanced their performance.

Conversely, a study by Hildebrandt et al. (2015) found AN patients had greater difficulty completing a PRL task (with feedback symbolized by a food-related image) compared to healthy controls. Other research has shown that AN patients made more perseverative errors in the Wisconsin Card Sorting Test (WCST) than the healthy controls (Aloi et al., 2015; Galimberti, Martoni, Cavallini, Erzegovesi, & Bellodi, 2012). However, some studies reported no performance difference in the WCST between the two groups (Perpiñá, Segura, & Sánchez-Reales, 2017; Tchanturia et al., 2012).

Roberts, Tchanturia, Stahl, Southgate, & Treasure (2007) conducted a comprehensive examination of various CI assessment tools employed in AN research, including the WCST, Trial Making Test, Brixton Spatial Anticipation Test, Haptic Illusion Task, and the CatBat. Most of these tools focused on the set/task-switching abilities of AN patients—the capability to shift attention from one set of stimuli to another. They found that AN patients generally struggled more than controls. However, this was challenged by Galimberti et al. (2012), who found no such set/task-switching difficulties in AN patients when using an Intra/extra-dimensional set-shifting task.

A small number of studies have delved into cognitive inflexibility in BN (Darcy et al., 2012; Roberts et al., 2007) and BED (Cury et al., 2020; Mobbs, Iglesias, Golay, & Van der Linden, 2011; Zhang, Manson, Schiller, & Levy, 2014). The existing research on these topics is limited and yields mixed results. While some researchers identified a set-shifting deficit in BN patients (Roberts et al., 2007), others found no differences in cognitive flexibility between BN patients and HC (Darcy et al., 2012). In the context of BED, Zhang et al. (2014) reported a CI deficit in tasks involving food-related feedback, yet other studies observed no notable differences in cognitive flexibility task performance between these patients and HC. Thus, there is a pressing need for additional studies in this area.

The research field on cognitive flexibility in eating disorders is mixed, with varied methodologies and outcomes. This underscores the need for further studies to gain a clearer understanding and perhaps achieve more consistent findings on the topic.

1.5 Cognitive Inflexibility in Eating Disorders: Open Questions

From this brief review, several insights emerge:

1) Most of the studies reviewed, primarily focus on general cognitive and cerebral dysfunctions as explanations for CF deficits. While this approach is valuable, ambiguities remain about the nature of CF deficits. As highlighted by Morris & Mansell (2018), it is debated whether this deficit represents a fundamental, consistent trait of an individual, termed a *domain-general* deficit, or if it manifests only in specific contexts, referred to as *domain-specific*. This perspective underscores the need to complement neuroscientific research with behavioral and contextual studies

to gain a comprehensive understanding of psychopathological dynamics. When researching CI in eating disorders, considering the nature of cognitive processes is crucial to ascertain the presence (or absence) of this deficit in eating disorders. Without this approach, interpretations may remain confused. Indeed, a cognitive process may function well under normal circumstances, but it can fail when faced with challenging contexts. Therefore, examining CI solely as a broad dysfunction might lead to ambiguous findings.

2) Previous studies have often categorized participants as either healthy controls or patients, thereby overlooking individual variations within these groups. Yet, significant disparities may exist even within these categories. As highlighted by Davis et al. (2020), patients can exhibit considerable differences, particularly in terms of symptom severity. Acknowledging these distinctions is crucial, as cognitive deficits might manifest differently based on the specific nature of symptoms. This variation has direct implications for treatment; strategies effective for patients with milder symptoms may not be suitable for those more severely affected. Therefore, an accurate assessment of cognitive deficits, which takes symptom severity into account, is essential for understanding the role of cognitive processes in both the onset and perpetuation of the disorder. Moreover, it is important to distinguish among different diagnostic categories, each presenting a variety of symptoms, especially in the realm of eating disorders, in order to comprehend cognitive processes like CI, considered *transdiagnostic*. In this context, comparing various eating disorders in the study of CI is crucial to elucidate the nature of this cognitive deficit.

3) Cognitive Flexibility (CF) is a multifaceted construct, with *set/task-switching* and *reversal learning* being its primary defining abilities. However, these two facets differ significantly, both neurally and behaviorally. Consequently, it is imperative to assess them with distinct tools tailored to each facet. Many previous studies overlooked this differentiation, leading to inconsistent results. The aforementioned literature displays an array of assessment instruments for CF, often without recognizing its multifaceted nature. A study by Hagan et al. (2020) underscores this disparity

in evaluation methods, emphasizing the importance of adopting a unified and standardized assessment approach.

To deepen our understanding of CI and its role on various psychopathologies, it is essential to address the existing theoretical challenges. These challenges are intrinsically connected to the methodologies used in research. In this vein, the current research project is designed to investigate CI as a critical *transdiagnostic* factor in mental disorders, with a primary focus on Eating Disorders, particularly AN.

Although focusing on AN is useful, it does not allow for a direct test of the *transdiagnostic* hypothesis of CI. Therefore, to test this hypothesis more directly, the project includes a comparative analysis with BN. This comparison is pivotal because it enables the examination of CI across different yet related disorders. By studying CI in both AN and BN, the research aims to determine whether CI is consistently present across these disorders, thereby supporting its role as a *transdiagnostic* factor.

Finding the presence of CI across different diagnoses, specifically AN and BN, would pave the way for extending these investigative methods to other psychological disorders. Such an expansion would be significant, as it could help establish CI as a common thread in various forms of psychopathology. In sum, the broader application of these methods will contribute to a more comprehensive understanding of the role of CI across the spectrum of mental health conditions.

Goals of the current study

The objective of this study is to clarify the role of CI in Eating Disorders, with a primary focus on AN and, to a lesser extent, BN. CI is proposed as a pervasive *transdiagnostic* factor, which is crucial for a comprehensive understanding of psychopathology and the development of effective therapeutic strategies.

To shed light on the subject, it is essential to address specific limitations identified in previous literature:

1) In the field of cognitive psychology, an ongoing but not yet fully resolved debate revolves around the nature of cognitive deficits. The core question focuses on whether these deficits should be considered domain-general, implying widespread cerebral dysfunctions affecting behavior across a wide range of situations, or conversely, as generally efficient mechanisms that only display limitations in particularly significant or salient contexts, referred to as domain-specific (Frensch & Buchner, 1999). This dilemma significantly influences the approach, understanding, and addressing of cognitive dysfunctions. Morris & Mansell (2018) has expanded this discussion by focusing on cognitive rigidity: is it an inherent trait of an individual or does it arise specifically in certain situations? In the case of AN, cognitive rigidity has traditionally been perceived as a pervasive feature of the disorder. However, this perspective has yielded inconsistent or conflicting evidence. It is essential to investigate whether the rigidity observed in AN patients is a general trait or primarily emerges in situations that trigger typical symptoms of the disorder. If confirmed, this could clarify the inconsistency in previous study results, suggesting that the construct may not have been adequately explored. This study aims to explore the

influence of context on the cognitive flexibility of individuals with AN. This will be achieved by using a Probabilistic Reversal Learning (PRL) task with ED-salient stimuli and a classic PRL task with neutral stimuli, followed by a comparison of the participants' performances in both scenarios (**Study 1**).

2) Eating disorders, especially AN, have been extensively researched. However, specific characteristics of the study samples are often neglected. Combining diverse patients with varying symptoms and severities into a single group can be misleading. For instance, patients with different types of AN (*e.g.*, restrictive type or binge-purging) and varying symptom severity may exhibit distinct cognitive profiles (Davis et al., 2020). Thus, it is crucial to account for these characteristics when investigating cognitive inflexibility or other cognitive deficits in AN. In this study, to assess the impact of symptom variability, the performance on a PRL task will be compared between a group of patients with severe AN symptoms and another with milder AN symptoms. The aim is to determine whether distinct cognitive profiles emerge within a population sharing the same diagnosis, yet exhibiting varying degrees of symptom severity. Furthermore, a group of BN patients will also be tested using a PRL task, to detect any potential cognitive flexibility deficits in accordance with the *transdiagnostic* perspective (**Study 2**).

3) Cognitive flexibility is a multifaceted and intricate construct. Historically, it has been linked to the concept of *set/task-switching.* Standardized tasks like the WCST, TMT, or Brixton Test are commonly used to assess this dimension. However, despite the widespread use of these tools, the literature presents inconsistent results concerning cognitive flexibility, especially in the context of eating disorders. Such discrepancies highlight the need to explore other components of cognitive flexibility. Recently, there has been growing academic interest in *reversal learning* as a critical dimension of cognitive flexibility. This facet offers a more dynamic perspective on CI, as highlighted by Hildebrandt et al. (2015). It is suggested that this dimension might be more closely associated with the persistence of symptoms in eating disorders than the traditional *set/task-switching* perspective. As such, this project will examine the

two facets separately within the context of AN, using specific tasks: the Probabilistic Reversal Learning for *reversal learning* and Task-Switching for *set/task-switching*. Additionally, the WCST will be employed, given its widespread recognition in the scientific community. The aim is to shed light on the roles of the two processes in AN by determining which one is impaired, while ensuring methodological rigor in the investigation (**Study 3**).

4) In the field of cognitive flexibility research, the wide range of tools employed poses a challenge when comparing results across different studies (Hagan et al., 2020). This complexity is further exacerbated as a single tool can be subjected to notably different data analysis techniques, as highlighted by Miles et al. (2021). Given this scenario, there is a clear need for a more standardized and consistent approach. Recently, computational models have been suggested as a promising approach for interpreting data from neuropsychological tasks (Weiss et al., 2021). Computational models present several advantages: (1) They monitor and analyze the learning process continuously throughout a task, offering an evolving framework to understand an individual's cognitive abilities; (2) rather than focusing on individual trials or isolated transitions between trials, computational models consider the entire dataset, facilitating a comprehensive analysis of performance; (3) these models generate a set of parameters that reflect various aspects of learning, allowing for a more nuanced understanding of the individual's performance beyond the mere sum of the feedback received during the task. While the computational approach holds immense potential, it is a relatively new approach in the field of psychology. It is important to acknowledge that these models do not yet apply universally to all types of tasks. Nevertheless, embracing computational modeling could represent a significant advancement toward a more standardized and insightful research methodology for assessing cognitive flexibility. Therefore, in all the studies conducted within this project, data will be analyzed using advanced computational models (**Study 1, Study 2, and Study 3**).

Chapter 2

Study 1: The Destructive Impact of Food-Related Information on Cognitive Flexibility in Anorexia Nervosa

2.1 Introduction

Anorexia Nervosa (AN) is a complex psychiatric disorder with significant consequences on physical and psychological health. However, its underlying mechanisms remain not fully understood. While recent studies have highlighted cognitive flexibility deficits and anomalies in feedback processing in AN, the role of context in influencing these factors remains unexplored. The present study deepens the understanding of how contextual elements can influence decision-making processes in individuals with AN.

2.1.1 Anorexia nervosa (AN): Characteristics and Demographics

AN is a severe and notoriously difficult to treat psychiatric disorder characterized by an excessive preoccupation with body weight and shape (American Psychiatric Association, 2022).

Patients with AN frequently drastically reduce their caloric intake and avoid specific foods. Such behaviors lead to significant weight loss and consequently to a marked loss of muscle mass and adipose tissue, particularly in the restrictive type (AN-R). Significant weight loss affects psycho-physical health, leading in the most severe cases to cardiovascular problems, damage to internal organs, osteoporosis, mood fluctuations and anxiety (American Psychiatric Association, 2022). Another key feature of AN is distorted body image. Individuals suffering from AN perceive themselves as overweight, even when critically underweight. The ability to maintain strict weight control determines the feeling of self-esteem and personal worth (Fairburn et al., 2003). Although traditionally associated with females, with a prevalence estimated at 1.4%, recent studies have also identified cases of AN in males, with a lower prevalence of 0.2% (Galmiche, Déchelotte, Lambert, & Tavolacci, 2019; Smink, Hoeken, & Hoek, 2013). Mortality rate associated with AN can be as high as 5-20% (Qian et al., 2022).

Despite the extensive body of scientific literature dedicated to AN, its etiology still remains a focal point of research. The lack of a comprehensive understanding of AN makes therapeutic intervention highly complex. With success rates falling below 50%, there is a high predisposition to relapse and frequent diagnostic transitions between various eating disorders (Atwood & Friedman, 2020; Linardon, Fairburn, Fitzsimmons-Craft, Wilfley, & Brennan, 2017).

Therefore, deepening our understanding of the underlying mechanisms that drive the onset and persistence of AN is essential in developing more effective therapeutic interventions (Chang, Delgadillo, & Waller, 2021).

2.1.2 Cognitive Flexibility and Reinforcement Learning in AN

Recent research posits that deficits in executive functions, specifically cognitive flexibility (*i.e.*, the ability to adjust behavior and thinking based on environmental feedback), play a crucial role in the onset and persistence of AN (Bartholdy, Dalton, O'Daly, Campbell, & Schmidt, 2016; Guillaume et al., 2015; Wu et al., 2014).

A cognitive flexibility deficit, manifesting as rigidity or inflexibility, can result in repetitive behaviors due to impaired reinforcement-based learning. Reinforcement learning is the mechanism by which humans learn from the consequences of their actions: positive feedback promotes certain behaviors, while negative outcomes discourage certain behaviors or similar ones, signaling the need for change. According to Walsh's model, in AN, pronounced rigidity hinders the acquisition of new adaptive responses, potentially leading to challenges in acknowledging the detrimental effects of restrictive eating, thereby maintaining the disorder (Walsh, 2013).

Anomalies in feedback processing are believed to underpin this mechanism (Schaefer & Steinglass, 2021).

Notably, individuals with AN often exhibit heightened sensitivity to perceived punishments and diminished responsiveness to rewards, particularly when they are exposed to stimuli that are salient to eating disorders (ED-salient stimuli), such as food-related stimuli (Wierenga et al., 2014). This distorted perception leads them to avoid situations they see as negative, hindering their ability to correctly interpret adverse events or stimuli. As a result, AN patients actively avoid what they perceive as punitive (Jonker, Glashouwer, & Jong, 2022; Matton, Goossens, Braet, & Vervaet, 2013).

Neuroimaging research corroborates these observations, revealing neural changes in AN patients related to their reactions to punishment and aversive food cues (Bischoff-Grethe et al., 2013; Monteleone et al., 2017; Wagner et al., 2007). In contrast, AN is associated with decreased reward sensitivity and a reduced neural reaction to stimuli typically considered gratifying.

Actions typically experienced as positive, such as consuming a favored food, might be seen as punitive or neutral by AN patients. Conversely, activities commonly viewed as negative might be perceived as rewarding, like severe food limitation or overexercising (Keating, 2010; Keating, Tilbrook, Rossell, Enticott, & Fitzgerald, 2012; Selby & Coniglio, 2020). Therefore, when situations require flexibility for an appropriate environmental response, such as selecting healthy foods or demonstrating acceptable social behaviors, their altered perception and diminished sense of reward might impede their ability to learn and adapt.

However, these are merely hypotheses, and the precise influence of these neural mechanisms on reinforcement learning in AN patients remains unclear.

Numerous studies have unsuccessfully attempted to clarify the role of altered reward processing in eating disorders (Bischoff-Grethe et al., 2013; Glashouwer, Bloot, Veenstra, Franken, & Jong, 2014; Harrison, Genders, Davies, Treasure, & Tchanturia, 2011; Jappe et al., 2011; Matton et al., 2013). For instance, Ritschel et al. (2017) using a traditional Probabilistic Reversal Learning (PRL) task with neutral stimuli found that patients with recovered-AN had compromised PRL performance when presented with negative feedback, compared to healthy controls. Conversely, Bernardoni et al. (2018b) discovered that AN patients displayed a superior learning rate in the PRL task but only in trials involving punishment. Sarrar et al. (2016) found no performance differences in the task between acute AN patients and healthy controls using the Probabilistic Object Reversal Task. Similarly, Geisler et al. (2017) found no group differences in a PRL task with neutral stimuli and monetary feedback.

The evidence underscores the complex nature of the problem and emphasizes the challenges in drawing definitive conclusions.

2.1.3 The Role of Context in Cognitive Flexibility: A Proposal

The aforementioned studies have not conclusively determined the presence or absence of a reinforcement learning deficit in AN. Specifically, some studies identified this deficit predominantly in response to negative feedback, while others reported enhanced performance in the presence of negative feedback. Additionally, some research found no differences in reinforcement learning capabilities between AN patients and control groups.

Previous studies employed a Reinforcement Learning (RL) task using stimuli and feedback without emotional significance, leading to conflicting conclusions. However, it is possible that alterations in reinforcement learning may not manifest in situations perceived as neutral. On the other hand, challenges in flexibly adapting to the environment may become more evident, or perhaps even exacerbated, in contexts seen as particularly relevant (*e.g.*, ED-salient circumstances).

This proposal shifts the emphasis from a domain-general to a domain-specific approach, suggesting that the cognitive flexibility deficit may not be an inherent trait of AN individuals, manifesting across all contexts. Instead, it may be a cognitive process that functions adequately but becomes disrupted when exposed to specific contextual elements (Hitchcock et al., 2022). This domain-specific approach challenges the conventional neuroscientific perspective on cognitive deficits.

Recently, several researchers have attempted to adopt the domain-specific perspective, integrating food-related cues into the RL paradigm (Smith, Mason, Johnson, Lavender, & Wonderlich, 2018). Zhang et al. (2014) administered two versions of a PRL task to a group of obese individuals without Binge Eating Disorder (BED), a group with BED, and a control group. While one version of the task presented monetary feedback, the other offered food-related feedback. Their findings indicated that those with BED struggled more with the task when confronted with food-related feedback, as opposed to monetary feedback. This implies that the presence of food-related cues may compromise reinforcement learning abilities.

However, replication attempts of these findings in AN have produced inconsistent results. Hildebrandt et al. (2015) documented increased rigidity in AN individuals using a PRL task with food-related feedback, whereas Hildebrandt et al. (2018) observed no performance differences between AN patients and healthy controls using the same paradigm. To delve deeper into the domain-specific nature of the cognitive flexibility deficit, it is imperative to manipulate the context surrounding decision-making. Solely modifying feedback is not comprehensive, as it accentuates the outcomes of decisions without addressing the broader contextual elements that influence those decisions (Schaefer & Steinglass, 2021).

Thus, investigating the effect of context, rather than merely feedback manipulation, on the adaptive capabilities of patients with Eating Disorders seems crucial. As highlighted by Smith et al. (2018), cognitive biases in these patients only manifest in the presence of ED-salient stimuli. For instance, patients with BED and Bulimia Nervosa (BN) experience inhibitory control deficits when confronted with body or food-related stimuli, yet no such deficits are evident with neutral stimuli (Kittel, Brauhardt, & Hilbert, 2015). Likewise, patients with AN, BN, and BED demonstrate attention biases during tasks such as the *Stroop Test* and *Dot-probe Test* that feature food-related stimuli. These attention biases are not observed in the same patients when neutral stimuli are presented (Brooks, Prince, Stahl, Campbell, & Treasure, 2011).

Despite these evidence, previous research on cognitive flexibility deficits in AN largely neglected the hypothesis that these impairments might be specific to ED-salient contexts (*i.e.*, domain-specific). However, it is imperative to discern whether the reinforcement learning deficit in AN is domain-specific, as opposed to suggesting a general learning process alteration (Haynos et al., 2022). Such a perspective could shed light on the inconsistencies found in prior research regarding the presence of a cognitive flexibility deficit in AN, since the domain-specificity hypothesis has not been adequately tested.

From a theoretical perspective, prior studies have potentially underestimated the significance of context in shaping flexible responses to environmental cues. In the present study, the emphasis is on the potential contextual influences that may compromise decision-making in AN patients.

In summary, this overview of the complex mechanisms underlying AN underscores the crucial importance of adopting a domain-specific perspective when investigating cognitive processes. It is hypothesized that cognitive flexibility deficits are not solely inherent traits but are also significantly influenced by environmental cues. By examining the contextual determinants, a deeper insight into decision-making in AN is obtained, emphasizing the importance of therapeutic interventions that take into account both internal and external factors.

2.2 The Present Study

Cognitive inflexibility in AN may not represent a universally compromised neural mechanism. Instead, this mechanism may operate effectively in specific contexts but fail in others. Notably, rigidity might be more pronounced in situations highly relevant to

AN patients, such as food-related contexts. This insight carries profound implications, indicating that treatments should target specific facets of the issue rather than resorting to a broad-based solution. It is crucial to understand why the mechanism fails in certain scenarios and find strategies to enhance its effectiveness in these challenging situations. However, a pivotal preliminary step is to verify solid evidence for domain-specificity in these deficits.

Consequently, the purpose of the present study is to examine the potential deficit in cognitive flexibility among patients diagnosed with Anorexia Nervosa-Restrictive subtype (AN-R). Specifically, the study aims to determine whether such a deficit is domain-specific or whether it represents a generalized impairment (*e.g.*, domain-general).

Cognitive inflexibility has been proposed as a potential underlying factor for both the maintenance and onset of AN-R. To evaluate its role in maintaining the disorder, a group of AN-R patients was examined. Conversely, to assess its contribution to the development of the disorder, a group of individuals at risk of developing eating disorders was tested. Being "at-risk" denotes displaying dysfunctional eating behaviors that could, over time, lead to the development of the disorder. Therefore, if cognitive inflexibility plays a role in the development of the disorder, it should be evident even in the premorbid phases among individuals who already exhibit features that make them at-risk.

To test the hypothesis, two versions of a PRL task were administered to a healthy control group (HC), a group at risk of developing an eating disorders (RI), and a group of patients with AN-R: a traditional version with neutral stimuli (*e.g.*, a flower and a book) and a modified version in which a neutral stimulus was paired with a food stimulus (*e.g.*, french fries).

Considering the alterations in basic cognitive processes commonly observed in individuals with an eating disorders when confronted with ED-salient stimuli (Haynos, Lavender, Nelson, Crow, & Peterson, 2020), it is plausible to hypothesize the following:

H1: Patients with AN-R will exhibit a diminished learning rate in the PRL block featuring ED-salient stimuli, but not in the neutral block. The same is expected for the at-risk group. On the other hand, no learning rate differences between the two blocks are expected for the control group.

H2: Patients with AN-R are expected to display a reduced learning rate in the PRL block involving ED-salient stimuli choices, compared to both the healthy controls and those at risk of developing an eating disorders.

H3: Patients with AN-R are anticipated to exhibit a slower task execution and delayed decision-making compared to both healthy controls and the at-risk group, especially in the domain-specific PRL block.

Lastly, no difference is anticipated between AN-R patients, the at-risk group, and healthy controls in the neutral PRL block, where choices are unrelated to ED-salient themes or contexts.

2.3 Method

2.3.1 Participants

The overall sample consisted of 40 outpatients diagnosed with Restrictive Anorexia Nervosa (AN-R), 45 healthy controls (HC), and 38 healthy individuals at risk of developing an eating disorder (RI), matched for age, gender, and education level (Table 2.1). All three groups comprised female participants. Patients were recruited from an Italian center specializing in the treatment of eating disorders, namely the Specchidacqua Institute in Montecatini (Pistoia, Italy). All recruited patients met the criteria for an Anorexia Nervosa-Restrictive subtype diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5, American Psychiatric Association, 2013). The diagnostic evaluation was carried out by professionals, psychiatrists and psychologists, affiliated with the aforementioned institute. Data collection was conducted approximately 2 weeks after patients were admitted and initial assessments were completed. In addition to the primary diagnosis of eating disorders, psychiatric comorbidities were also considered. These were identified during the diagnostic process through additional psychological evaluations conducted by the mental health professionals from the institute. Additionally, the medication status of each patient was taken into account. The same professionals assessed the patients' research participation eligibility criteria using semi-structured clinical interviews. Exclusion criteria included the presence of neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis.

The HC group was voluntarily recruited through advertisement on social media or at the university. Subsequently, HCs underwent the screening test Eating Attitudes Test-26 (EAT-26) to assess potential tendencies toward developing an ED. Participants scoring above the cut-off ($\geq$ 20) on the EAT-26 were considered at risk of developing eating disorders and classified as "at-risk" individuals (Dotti & Lazzari, 1998). None of the HC participants reported currently undergoing treatment or having a previous ED diagnosis. Both the HC and "at-risk" individuals were evaluated against the inclusion criteria to ensure eligibility. Exclusion criteria encompassed abnormal Body Mass Index (BMI) values, neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis.

All participants reported normal cognitive functioning, as assessed by the Raven's Standard Progressive Matrices (Raven, 2003). The majority of participants were Caucasian. All participants were right-handed, demonstrated proficiency in reading/writing and understanding the Italian language, and were kept unaware about the purposes of the study.

2.3.2 Materials

Self-report Measures

The *Eating Attitude Test-26* (EAT-26, Garner, Olmsted, Bohr, & Garfinkel, 1982) is a self-report scale consisting of 26 items. The EAT-26 evaluates potential disordered eating behaviors over the past three months through three subscales: the *Dieting* scale (*e.g.*, "I engage in dieting programs"), the *Bulimia and Food Preoccupation* scale (*e.g.*, "I feel an urge to vomit after meals") and the *Oral Control* scale (*e.g.*, "I avoid sweet foods"). A total score of $\geq$ 20 indicates the presence of potentially risky eating behaviors for the development of an eating disorder. Participants evaluated the intensity associated with each questionnaire item on a 6-point Likert scale (0 = Never, Rarely, Sometimes; 3 = Always). The EAT-26 is the most commonly used instrument in studies on ED in non-clinical populations (Dotti & Lazzari, 1998). The Italian version of the EAT-26 (Dotti & Lazzari, 1998) showed good psychometric properties (Cronbach's alpha = 0.86). Furthermore, Cronbach's alpha was found to be high for each subscale in a university sample (*Dieting* = 0.87; *Bulimia and Food Preoccupation* = 0.70; *Oral Control* = 0.62).

The *Body Shape Questionnaire-14* (BSQ-14, Dowson & Henderson, 2001) includes 14 items assessing overall body satisfaction over the past two weeks (*e.g.*, "I was ashamed of my body"). Intensity of concerns about physical appearance is rated on a 6-point Likert scale (1 = Never, 6 = Always). High scores indicate greater body dissatisfaction. The Italian version of the BSQ-14 has shown acceptable psychometric properties (Cronbach's alpha = 0.93) and good convergent validity (Matera, Nerini, & Stefanile, 2013). In the current sample, $\omega = 0.978$.

The *Social Interaction Anxiety Scale* (SIAS, Mattick & Clarke, 1998) assesses the presence of social anxiety through 20 items (*e.g.,* "In social situations, I usually feel uncomfortable"). Participants rated the intensity associated with the items on a 4-point Likert scale ranging from 0 (Not at all true) to 4 (Extremely true). Both the original and the Italian versions (Sica, Musoni, Bisi, Lolli, & Sighinolfi, 2007) demonstrate good psychometric properties (in the current sample $\omega = 0.938$).

The *Depression Anxiety Stress Scale-21* (DASS-21, Lovibond & Lovibond, 1995) consists of 21 items assessing symptoms of depression (*e.g.,* "I couldn't feel any positive emotions"), anxiety (*e.g.,* "I felt a lot of tension and found it hard to relax"), and stress (*e.g.,* "I felt stressed") over the past week. Items are rated on a 4-point Likert scale ranging from 0 (It never happened to me) to 3 (It happened to me almost always). Both the original and its Italian adaptation (Bottesi et al., 2015) demonstrate adequate reliability. In the current sample, $\omega_{\text{anxiety}} = 0.875$, $\omega_{\text{depression}} = 0.914$, $\omega_{\text{stress}} = 0.899$; for the total scale, $\omega = 0.945$.

The *Rosenberg Self-Esteem Scale* (RSES, Rosenberg, 1965) measures individual self-esteem levels using a 10-item scale (*e.g.,* "I feel that I'm a person of worth, at least on an equal basis with others"). Each item is rated on a 4-point Likert scale ranging from 4 (Strongly Agree) to 1 (Strongly Disagree). High values indicate good self-esteem.

In the current sample, $\omega = 0.949$.

The *Frost-Multidimensional Perfectionism Scale* (F-MPS, Frost, Marten, Lahart, & Rosenblate, 1990) is a scale with 35 items that evaluates tendencies toward perfectionism. In its initial version, the F-MPS comprised six subscales. However, it was noted that such a division caused psychometric instability. Therefore, according to Stöber (1998), the F-MPS is characterized by four underlying factors: *Concerns over Mistakes and Doubts* (CMD) (*e.g.,* "Making mistakes would make me feel uncomfortable"), *Parental Expectations and Criticism* (PEC) (*e.g.,* "My parents set very high standards for me"), *Personal Standards* (PS) (*e.g.,* "If I did not set the highest standards for myself, I would be likely to end up a second-rate person"), and *Organization* (OR) (*e.g.,* "Organization is very important to me"). The original F-MPS, as well as its Italian adaptation (Lombardo, 2008), both exhibit adequate reliability. In the current sample, $\omega_{\mathrm{CMD}} = 0.919$, $\omega_{\mathrm{PS}} = 0.851$, $\omega_{\mathrm{PEPC}} = 0.946$, $\omega_{\mathrm{OR}} = 0.931$; for the total scale, $\omega = 0.932$.

Probabilistic Reversal Learning (PRL) Task

The *Probabilistic Reversal Learning* (PRL) is a Reinforcement Learning (RL) task that assesses the ability to flexibly modify behavior when an environmental change occurs, as indicated by positive and negative feedback (den Ouden et al., 2013). The PRL is considered the most precise task for measuring cognitive flexibility in a laboratory setting (Fradkin, Strauss, Pereg, & Huppert, 2018). The task consists of 160 trials. In each trial two stimuli are concurrently presented on the left and right sides of the screen (Figure 2.1). Both stimuli are associated with a specific probability of reward: one stimulus is rewarded in most trials (70%), while the other is rewarded in fewer trials (30%). The reward is symbolized by an image of a symbolic euro coin, whereas the punishment is represented by an image of a crossed-out symbolic euro coin. The participant's aim is to accumulate as many rewards as possible, maximizing their symbolic gain. Within a 3-second interval, the participant must select one of the two stimuli, discern the underlying task rule (*i.e.*, which of the two stimuli provides positive feedback most of the time, and choose it as often as possible). However, after a certain number of trials, the rule reverses, and the reward probabilities switch between the two stimuli. The stimulus that was previously rewarded in 70% of the trials will, after the rule reversal, be the one rewarded in 30%, and vice versa. Consequently, participants must flexibly adapt their choice behavior to continue maximizing symbolic gain. In this version of the task (low volatility), the rule reversal occurs three times, every 40 trials. The task is characterized by four epochs, within which the reward probabilities associated with the stimuli remain consistent (Figure 2.1). Participants were informed about the stimulus-reward contingency changes but were not given specific cues about how or when these changes would occur. At the end of the task, the total (symbolic) gain achieved was displayed. In the present study, participants were administered two blocks of the PRL, each consisting of 160 trials. The two blocks differed in the type of stimuli presented. Within one block, a food-related image (*e.g.*, a slice of cake) was displayed alongside a food-unrelated image (*e.g.*, a book) to test the contextual-influence hypothesis of the cognitive flexibility deficit in AN-R patients. On the other hand, in the other block only food-unrelated images were displayed. In the domain-specific block, the two images (food and neutral) were randomly selected from a set of food-related and food-unrelated images. All images used in the study were selected from the *International Affective Picture System* (IAPS) database (Lang & Bradley, 2007). The food category included images of french fries, cakes, pancakes, cheeseburgers, and cupcakes (IAPS #7461, 7260, 7470, 7451, 7405). In contrast, the non-food category featured images of a lamp, book, umbrella, basket, and clothespin (IAPS #7175, 7090, 7150, 7041, 7052). For the neutral control task, five images were used for each food-unrelated stimulus: five images of flowers (IAPS #5000, 5001, 5020, 5030, 5202) and five images of objects (IAPS #7010, 7020, 7034, 7056, 7170). The experiment was controlled by the online platform Psytoolkit (https://www.psytoolkit.org).

2.3.3 Procedure

The data collection involved three phases. In the initial phase, inclusion and exclusion criteria for participants in the HC group were assessed. Those who voluntarily chose to participate in the study as control subjects were asked to provide their weight and

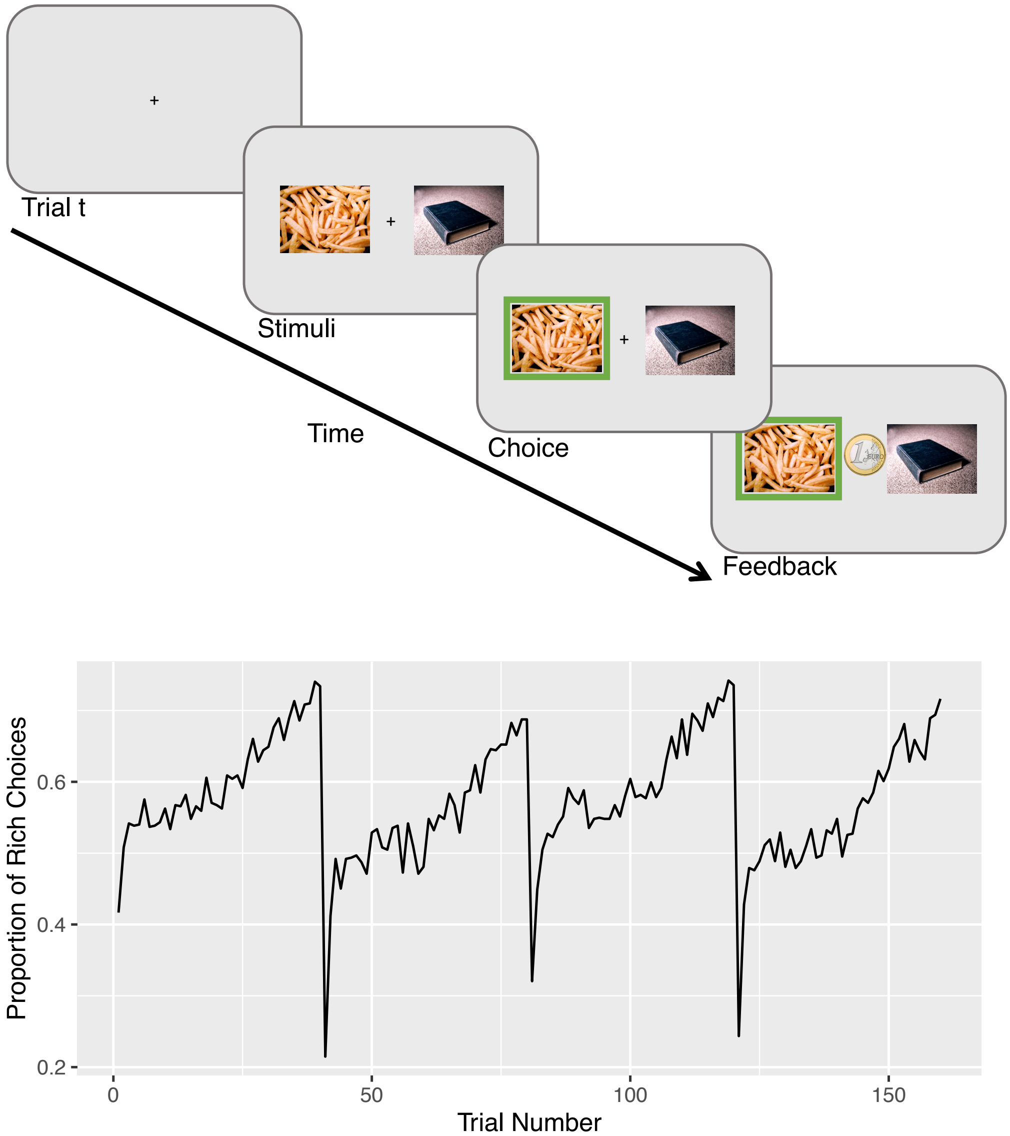

Figure 2.1: **Top.** Trial structure of the PRL task, from the display of the two stimuli to the presentation of feedback by choosing a certain stimulus. **Bottom.** the trial-by-trial proportion of the stimulus with the highest probability of reward, in each epoch.

height to calculate the BMI, and to complete the EAT-26 screening test. Participants scoring above the cut-off (≥ 20) were categorized as individuals "at risk" of developing an eating disorder. Participants scoring below the cut-off were classified as belonging to the HC group. Subsequently, a further evaluation was conducted to exclude the presence of neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis.

The patient group underwent similar screening procedures, administered by mental health professionals associated with the aforementioned recruiting institution.

In the second phase, all participants completed the Raven's Standard Progressive Matrices (Raven, 2003), aiming to evaluate their overall cognitive functioning. In addition to cognitive evaluations, participants filled out the six psychometric tests listed above.

Finally, in the third phase, participants completed the two blocks of the Probabilistic Reversal Learning Task. Each phase took about 30 minutes on average. Participation in the study was voluntary, and no incentives of any kind were provided to the participants. The study was conducted in accordance with the Declaration of Helsinki, and the procedure was approved by the Ethical Research Committee of the University of Florence (Prot. n. 0178082). All participants provided informed consent, received the privacy policy, and agreed to participate in the study.

2.4 Data Analysis

The PRL data were analyzed by estimating the six parameters of the computational model *Hierarchical Reinforcement Learning Drift Diffusion Model* (HDDMrl, Pedersen & Frank, 2020). The model delineates the learning trajectory across the PRL trials. Through computational analysis, the HDDMrl model decomposes the decision-making process into distinct components. This allows for insights into the underlying mechanisms of the learning process rather than solely relying on the overall task outcome as the only indicator of performance quality. To apply the model to the data, a Bayesian approach is required, since the estimation of computational models is currently feasible only through the Monte Carlo Markov Chain (MCMC) procedure. The Bayesian approach prioritizes estimating predicted values over hypothesis testing (Kruschke & Liddell, 2018). Therefore, for interpreting results, credible effects are considered by examining 95% credibility intervals or evaluating the proportion of posterior samples (97.5%) that indicate the direction of the effect.

Data analysis was performed using the Python software (https://www.python.org) and R version 4.2.1 (https://cran.r-project.org/index.html).

The HDDMrl allows to investigate the impact of ED-salient cues on decision-making. The model originates from the integration of two well-known computational models: The Rescorla-Wagner model (RW, Rescorla & Wagner, 1972) and the Drift Diffusion Model (DDM, Ratcliff & McKoon, 2008).

Rescorla-Wagner Model (RW)

The Rescorla-Wagner model (RW) is a famous model in associative learning research (Rescorla & Wagner, 1972). It explain how the strength of associations between stimuli and outcomes evolve based on prediction errors (*i.e.*, the differences between expected and actual outcomes). When an outcome aligns with expectations, the association remains stable. Conversely, when a mismatch between expectation and outcome occurs, the association can either intensify or diminish, contingent on the nature of the discrepancy.

The formula that captures this is:

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

Where Q refers to the expected values for option a on trial i, I represents the actual reward (with values 1 or 0), and α is the learning rate, which scales the update of the expected value based on the difference between the expected and actual rewards. Due to its pivotal role in adjusting expectations, α is regarded as a fundamental parameter of the model.

Moreover, the RW model incorporates a softmax function, which describes how individual utilize the expected values of stimuli to make decisions. Within this softmax function, the probability of selecting option a is determined by its expected value in relation to other available options n, and this relationship is moderated 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 summary, the RW model determines two meaningful parameters:

- The learning rate (α), which represents the rate at which the expected reward of stimuli is updated based on prior feedback. A high α value signifies rapid updating of expectations, while a low α indicates a more gradual adjustment. This parameter can be differentiated for reward-based trials (α^{+}) and punished-based trials (α^{-}). Specifically, α^{+} is determined exclusively from trials that received a reward, while α^{-} is derived solely from trials that incurred a punishment.

- The inverse temperature parameter β captures the trade-off between exploration and exploitation. Specifically, it measures the propensity to choose stimuli either randomly (exploration) or based on maximizing potential gains (exploitation). A high value of β suggests a predominant exploitation strategy, whereas a low value indicates a greater inclination towards exploration.

Drift Diffusion Model (DDM)

The Drift Diffusion Model (DDM) is a well-established model in cognitive science, particularly used to explain decision-making processes in two-choice tasks. It assumes that decisions are made by accumulating noisy evidence over time until a certain threshold is reached (Ratcliff & McKoon, 2008).

Fundamentally, the DDM converts behavioral data, in terms of accuracy, mean response times, and the distribution of response times, into elements of cognitive processing. This allows for a granular understanding of how decisions evolve and are finalized over time.

The DDM is often mathematically represented by a stochastic differential equation that describes the accumulation of evidence over time. In its most basic form, the model can be written as follows:

$$dX_t = v \cdot dt + \sigma \cdot dW_t$$

Where dX_t is the accumulated evidence at time t, v is the *drift rate*, σ is the *noise* parameter, representing the standard deviation of the evidence accumulation process, which captures the random nature of evidence accumulation, and dW_t is a Wiener process (or Brownian motion), which is a stochastic process that models the random fluctuations in the evidence accumulation process.

The decision is made when the evidence X_t reaches one of the two boundaries (thresholds), usually denoted as a and $-a$ for a two-choice task.

In summary, the DDM model determines four meaningful parameters:

- The *drift rate* (v) represents the average rate at which evidence accumulates during the cognitive processing of stimuli. It reflects the predisposition toward either swift and accurate decisions or slower, more cautious ones. A high v indicates faster cognitive processing and a clear distinction between choices, while a low v suggests more uncertainty or slower processing.

- The *decision threshold* (a) quantifies the amount of evidence needed to reach a decision. A high a indicates a preference for more deliberate, cautious decision-making, requiring more evidence before deciding. Conversely, a low a suggests impulsiveness, where decisions are made with less evidence.

- The *non-decision time* (t) measures the combined time taken for processes outside the decision-making itself, such as perceiving the stimulus and initiating a motor response (*i.e.*, the interval from stimulus presentation to choice execution). A high t suggests longer periods dedicated to non-decision processes, whereas a low t indicates faster perception and action initiation.

- The *starting-point bias* (z) reveals any initial preference for one stimulus over another prior to feedback.

2.4.1 The Hierarchical Reinforcement Learning Drift Diffusion (HDDMrl) Computational Model

The HDDMrl model integrates parameters from both the RW and DDM models (Pedersen & Frank, 2020). From the RW model, it preserved the α parameter, which can be separately estimated for rewards and punishments. However, the β parameter from the RW model is replaced with parameters from the DDM. Since the β parameter, although useful in capturing the balance between exploratory and advantageous choices, does not delineate the cognitive and motor processes involved in decision-making and cannot distinguish between fast yet accurate decisions and those that are slow and cautious. On the other hand, sequential sampling models like the DDM encapsulate the dynamics of the decision-making process and their influence on the distribution of response times (Pedersen & Frank, 2020).

As a result, the HDDMrl model estimates six key parameters: - *Positive learning rate* ($alpha^+$) - *Negative learning rate* ($alpha^-$) - *Drift rate* (v) - *Decision threshold* (a) - *Non-decision time* (t) - *Starting point bias* (z).

Parameters were separately estimated for each group (AN-R, RI, HC) and for both PRL blocks (neutral and domain-specific). This allowed to examine how model parameters varied in response to different contextual conditions across the three groups.

These separately computed parameters were then analyzed in two distinct ways:

Within-group Parameter Comparisons. This approach contrasted parameter values between the neutral and domain-specific blocks within each group, capturing the effect of experimental manipulation within subjects.

Between-group Parameter Comparisons. This method contrasted the AN-R patient group with the group at potential risk for eating disorders, and with the control group. The comparisons were made distinctively for both the neutral and domain-specific blocks.

2.5 Results

Table 2.1 presents the descriptive statistics for the 3 participant groups. Specifically, the table displays core discrepancies sourced from the self-report questionnaires across the AN-R, HC, and RI groups. The data suggests that the AN-R patient group exhibits higher scores in the evaluated dimensions (anxiety, depression, stress, dysfunctional eating behaviors, low self-esteem, perfectionism, social anxiety, low BMI) compared to both the healthy controls and the at-risk subjects. However, individuals at risk of developing an ED demonstrate some characteristics in common with the AN-R patient group. At-risk subjects report levels of anxiety, depression, stress, self-esteem, social anxiety, and dysfunctional eating behaviors similar to those of the AN-R patient group. This suggests that even in a premorbid stage, individuals vulnerable to developing a particular disorder might exhibit characteristics closely mirroring those of individuals diagnosed with the disorder (Caudek et al., 2021; Pringle, Harmer, & Cooper, 2010). It is important to emphasize that the administered tests do not possess diagnostic value but serve descriptive purposes. The EAT-26 test, largely used with non-clinical populations (Dotti & Lazzari, 1998), is the only test capable of detecting the presence or absence of dysfunctional eating tendencies, suggesting a potential predisposition and susceptibility towards developing an ED.

Table 2.1: Demographic and Clinical Characteristics of the Sample.

	AN-R (n = 40) Mean (SD)	HC (n = 45) Mean (SD)	RI (n = 38) Mean (SD)	AN - HC PE (95% CI)	AN - RI PE (95% CI)	AN - HC Cohen's d	AN - RI Cohen's d
Age (years)	21.11 (4.33)	19.49 (2.32)	21.31 (4.82)	-0.00 (-0.22, 0.18)	0.01 (-0.21, 0.25)	0.09	-0.18
Education (years)	14.53 (1.11)	14.11 (0.98)	13.89 (0.78)	0.08 (-0.14, 0.33)	0.13 (-0.11, 0.41)	-0.08	-0.13
BMI (kg/m^2)	17.79 (2.85)	21.78 (3.53)	21.64 (4.11)	-2.74 (-3.67, -1.79)	-3.22 (-4.32, -2.07)	2.67	3.15
EAT-26 Total score	35.89 (19.46)	5.09 (5.10)	25.86 (12.44)	1.76 (1.32, 2.20)	0.18 (-0.25, 0.62)	-1.84	-0.18
EAT-26 Dieting	19.11 (11.24)	3.27 (3.96)	16.06 (8.00)	53.96 (32.78, 80.9)	4.38 (-2.46, 12.10)	-4.54	-0.36
EAT-26 Bulimia	7.17 (3.95)	0.76 (1.48)	5.78 (3.86)	1.36 (0.99, 1.72)	0.11 (-0.17, 0.41)	-2.33	-0.18
EAT-26 Oral control	9.61 (6.23)	1.07 (1.67)	4.03 (4.05)	1.53 (1.12, 1.96)	0.81 (0.42, 1.20)	-1.95	-1.03
BSQ-14 Total score	139.78 (35.26)	97.47 (32.37)	147.94 (37.13)	17.50 (11.30, 23.91)	-3.40 (-10.10, 3.15)	-1.21	0.24
RSES Total score	22.69 (5.29)	28.33 (5.76)	22.53 (5.76)	-5.601 (-8.12, -3.15)	0.192 (-2.31, 2.90)	-1.22	0.23
DASS-21 Stress	12.86 (4.67)	9.13 (3.55)	12.17 (3.74)	3.73 (1.89, 5.38)	0.69 (-1.12, 2.59)	-0.94	-0.18
DASS-21 Depression	10.61 (5.49)	6.82 (4.33)	11.22 (4.99)	3.78 (1.76, 6.14)	-0.62 (-2.91, 1.77)	-0.77	0.12
DASS-21 Anxiety	8.25 (4.51)	5.76 (4.26)	7.56 (4.47)	2.48 (0.50, 4.33)	0.70 (-1.39, 2.64)	-0.56	-0.16
SIAS Total score	37.31 (15.45)	27.69 (13.01)	39.03 (14.87)	9.63 (3.12, 15.84)	-1.68 (-8.28, 5.07)	-0.66	0.12
MPS CMD	45.47 (8.21)	40.02 (7.24)	49.25 (8.17)	5.46 (2.05, 9.02)	-3.73 (-7.47, -0.15)	-0.69	0.48
MPS PS	25.33 (5.74)	22.00 (4.86)	25.67 (6.33)	3.33 (0.89, 5.86)	-0.32 (-3.07, 2.19)	-0.60	0.06
MPS PEC	20.78 (6.64)	21.02 (5.84)	25.22 (7.92)	-0.63 (-3.49, 2.42)	-3.88 (-7.09, -0.90)	0.09	0.57
MPS OR	23.94 (5.11)	23.07 (5.16)	22.17 (5.55)	0.84 (-1.49, 3.17)	1.77 (-0.66, 4.26)	-0.17	-0.34

2.5.1 Quality Control

Before performing any statistical analysis, a quality control was applied to the PRL data. This quality check involves identifying those who performed the PRL task randomly, without following a rule that would allow them to make choices above chance level. These subjects were excluded from further analysis. The participants who met the quality control criteria (reaction times $\geq$ 150 ms, $\leq$ 2500 ms; accuracy below 40%, missing responses) and who were included in subsequent analyses were a total of N = 117, specifically N = 36 individuals with AN-R, N = 45 HC, and N = 36 RI.

2.5.2 Model Selection

The HDDMrl model can be implemented with varying degrees of complexity, ranging from the base model in which only standard parameters (α, a, v, t) are estimated, to the most complex model where all possible parameters are estimated. The most complex model estimates all parameters, differentiating among all the variables of interest (groups and type of stimulus presented). Selecting the most suitable model for the data is crucial in order to obtain the most accurate results possible. The *Deviance Information Criterion* (DIC) was used as the criterion to select the most accurate model. This criterion helps identify the model that achieves an optimal balance between the model fit to observed data and its intrinsic complexity. In practice, the model with the lowest DIC value is chosen, as it represents the best trade-off between data description precision and model complexity. In this study, the best trade-off was found to be the model M7, which estimates all possible parameters (α^-, α^+, a, v, t) separately for each group (AN-R, HC, RI) and for the type of task (neutral vs. food), with the exception of the z parameter related to response bias, which was found to be non-informative (for details, see Appendix A).

2.5.3 Results of the HDDMrl Model

The HDDMrl Model (M7) was estimated using 15,000 iterations, with a *burn-in* of 5,000 iterations. The $\hat{R}$ values for all estimated parameters were below 1.1, indicating that the model achieved good convergence.

To evaluate the influence of ED-salient information (*i.e.*, the domain-specific perspective) on cognitive flexibility in patients with AN-R as compared to the HC group and the RI group, two types of comparisons were conducted: (1) in the first comparison, the posterior estimates of the HDDMrl model parameters in both the domain-general and domain-specific conditions of the PRL were examined for each group separately. This comparison aimed to analyze within-group differences, elucidating how the conditions affect cognitive flexibility within each individual group; (2) the posterior estimates of the HDDMrl model parameters across all three groups were examined, considering both the domain-general and domain-specific conditions. This comparison aimed to explore between-group differences, investigating how cognitive flexibility varies among the groups under different conditions.

This approach was designed to provide insight into the specific impact of ED-salient information on cognitive flexibility in patients with AN-R and how this impact compares to that observed in the HC and RI groups under similar conditions.

Within-groups comparison of estimated parameters

First, we examine the comparison between the domain-specific block and the domain-general block within the three groups (Table 2.2). Participants in the AN-R group reported a lower learning rate on rewarded trials (α^+) for choices in the domain-specific block with food stimuli, compared to choices made in the neutral block, with stimuli not related to the disorder ($p = 0.0098$, Cohen d = -1.206) - Figure 2.2. No credible difference was observed in the learning rate for rewarded trials between the domain-specific block and the neutral block in the HC group ($p = 0.5544$) or RI ($p = 0.2247$) (Figure 2.3; Figure 2.4). No credible difference was observed in the learning rate for punished trials between food-related choices and food-unrelated choices for any of the three groups, AN-R ($p = 0.2349$), HC ($p = 0.6993$) and RI ($p = 0.5101$). Furthermore, it emerged that individuals with AN-R displayed a higher decision threshold (a) for choices evoking symptoms compared to neutral choices ($p = 0.0013$, Cohen's $d = 0.802$), indicating a more conservative tendency in decision-making (Figure 2.5). Similarly, though with a weaker effect, HC also exhibit a more conservative tendency in their choices in the block with food stimuli compared to the neutral block ($p = 0.0256$, Cohen's $d = 0.474$) - Figure 2.6. No credible difference was observed in the RI group for the parameter a ($p = 0.1026$) - Figure 2.7.

Table 2.2: Posterior estimates of the HDDMrl Model parameters by Group (AN-R, HC, RI) and PRL block (domain-specific vs. domain-general). The learning rates (α) are displayed on a logit scale. The probability (p) describes the Bayesian test comparing the posterior estimate of the parameter in the context of ED-salient information with that in the context of neutral information. Standard deviations are provided in parentheses.

Group	Parameter	Neutral choice	Food choice	p	Cohen's d
AN-R	a	1.273 (0.039)	1.442 (0.040)	0.0013	0.802
AN-R	v	1.403 (0.320)	1.776 (0.342)	0.7907	0.190
AN-R	t	0.188 (0.011)	0.174 (0.011)	0.8311	-0.253
AN-R	α^-	1.815 (1.081)	0.738 (1.096)	0.2349	-0.432
AN-R	α^+	1.006 (0.899)	-1.786 (0.756)	0.0098	-1.206
HC	a	1.222 (0.033)	1.314 (0.034)	0.0256	0.474
HC	v	2.157 (0.265)	1.790 (0.263)	0.1606	-0.358
HC	t	0.183 (0.009)	0.172 (0.009)	0.8228	-0.280
HC	α^-	2.780 (0.874)	3.442 (0.980)	0.6993	0.298
HC	α^+	1.198 (0.680)	1.326 (0.700)	0.5544	0.071
RI	a	1.245 (0.041)	1.316 (0.039)	0.1026	0.403
RI	v	2.197 (0.322)	1.849 (0.307)	0.2133	-0.381
RI	t	0.188 (0.011)	0.186 (0.011)	0.5462	0.166
RI	α^-	2.857 (1.067)	2.904 (1.062)	0.5101	0.015
RI	α^+	1.573 (0.847)	0.739 (0.752)	0.2247	-0.438

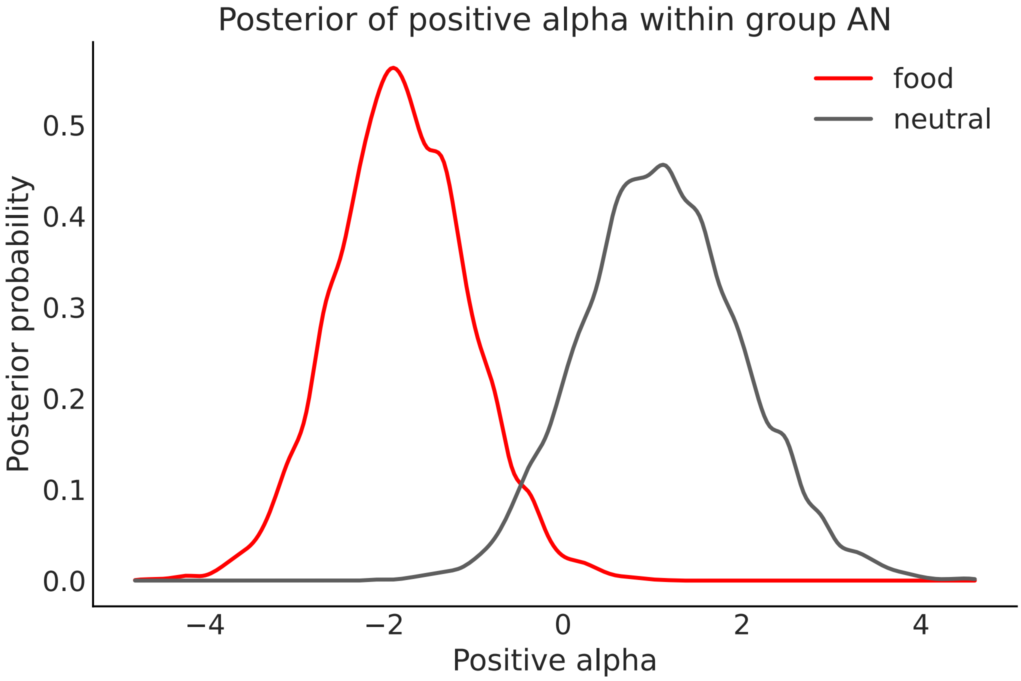

Figure 2.2: Comparison of the posterior distributions of the parameter α^+ estimated separately for the two PRL blocks (domain-general and domain-specific) in the AN group. The red curve denotes the posterior distribution for food-related trials, while the grey curve represents that for neutral trials.

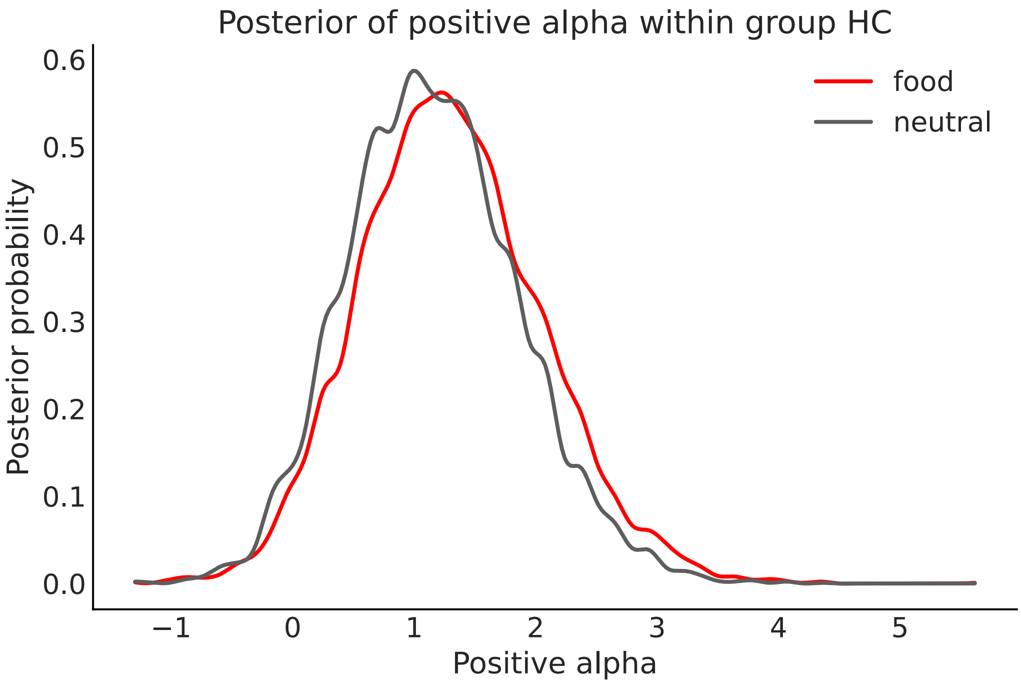

Figure 2.3: Comparison of the posterior distributions of the parameter α^+ estimated separately for the two PRL blocks (domain-general and domain-specific) in the HC group. The red curve denotes the posterior distribution for food-related trials, while the grey curve represents that for neutral trials.

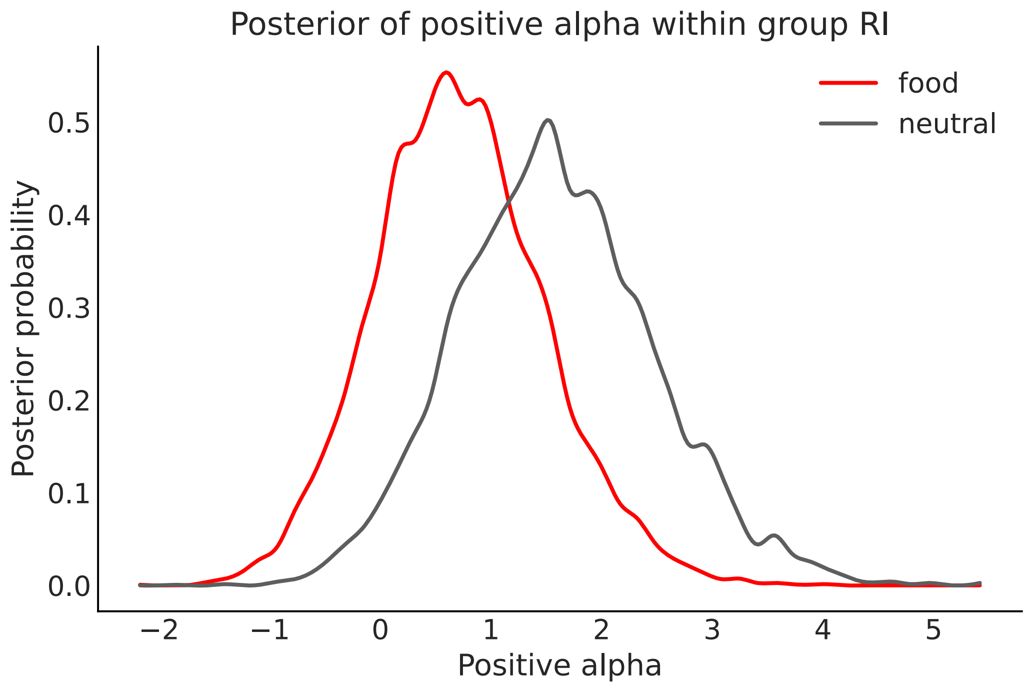

Figure 2.4: Comparison of the posterior distributions of the parameter α^{+} estimated separately for the two PRL blocks (domain-general and domain-specific) in the RI group. The red curve denotes the posterior distribution for food-related trials, while the grey curve represents that for neutral trials.

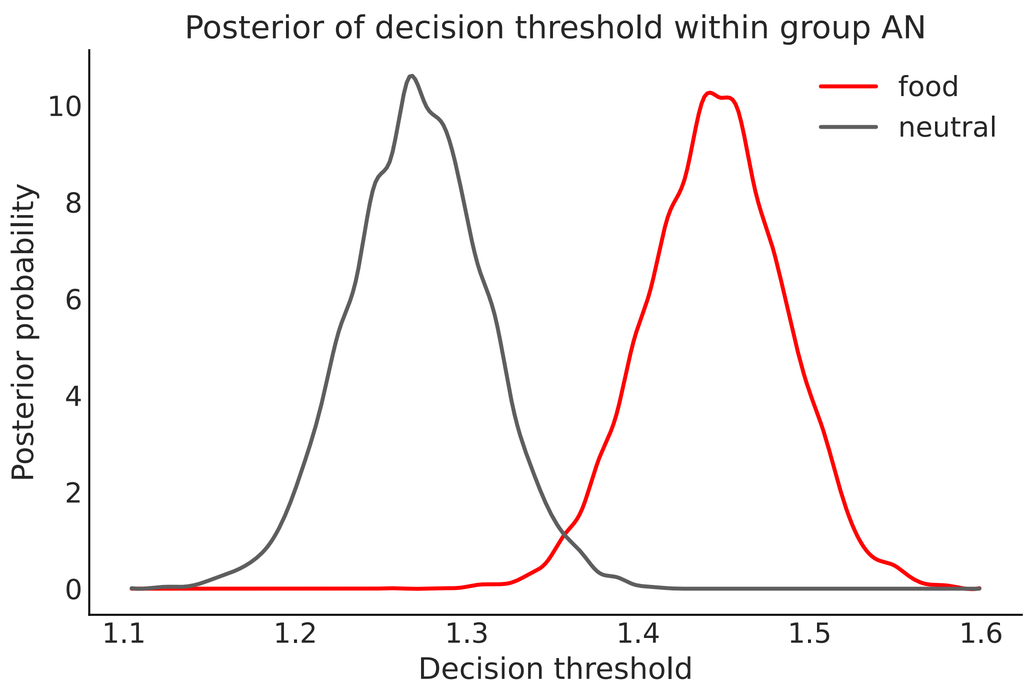

Figure 2.5: Comparison of the posterior distributions of the parameter a estimated separately for the two PRL blocks (domain-general and domain-specific) in the AN group. The red curve denotes the posterior distribution for food-related trials, while the grey curve represents that for neutral trials.

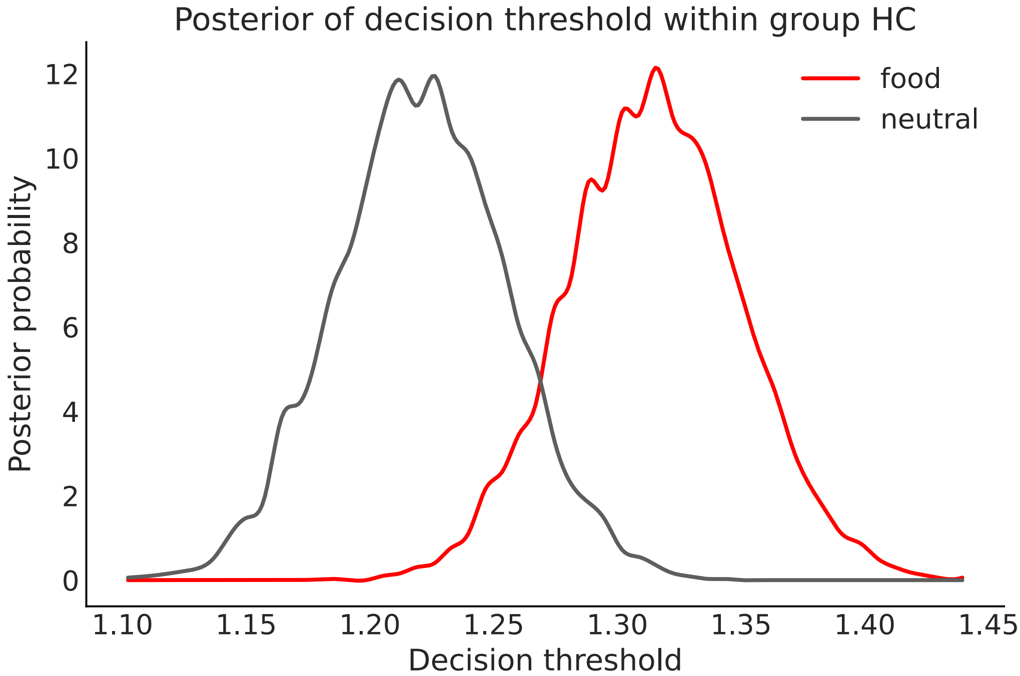

Figure 2.6: Comparison of the posterior distributions of the parameter a estimated separately for the two PRL blocks (domain-general and domain-specific) in the HC group. The red curve denotes the posterior distribution for food-related trials, while the grey curve represents that for neutral trials.

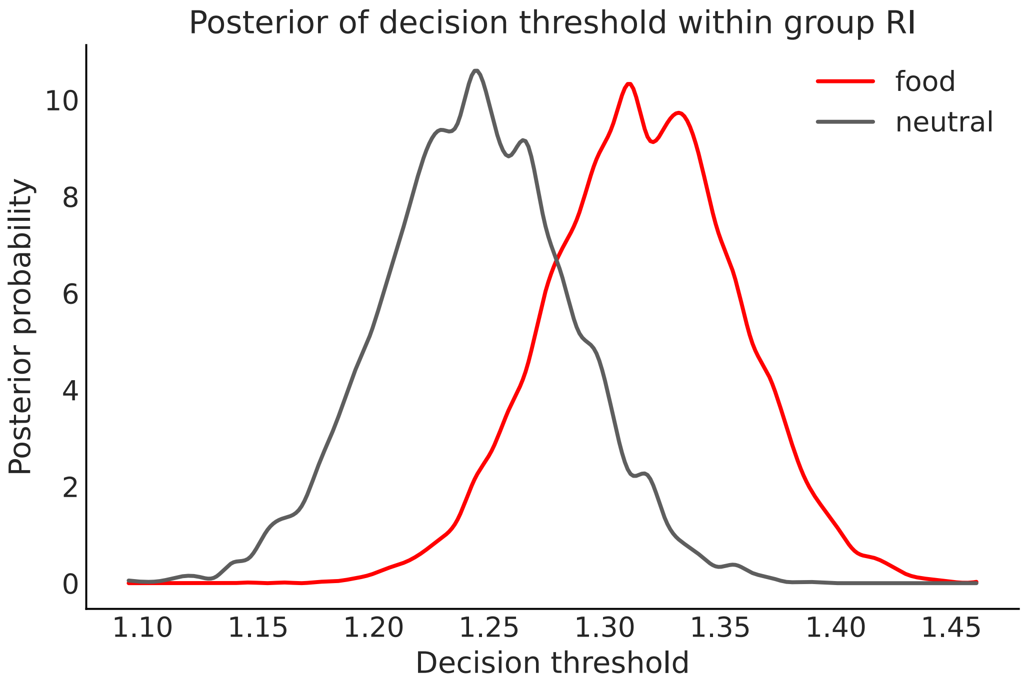

Figure 2.7: Comparison of the posterior distributions of the parameter a estimated separately for the two PRL blocks (domain-general and domain-specific) in the RI group. The red curve denotes the posterior distribution for food-related trials, while the grey curve represents that for neutral trials.

Between-groups comparison of estimated parameters

Here, we examine the groups differences in their performance in the PRL domain-specific block and the neutral block. In the domain-specific block, the AN-R group exhibited a reduced learning rate in rewarded trials compared to both the controls (p = 0.0009, Cohen's d = 1.498) and the at-risk individuals (p = 0.0085, Cohen's d = 1.209) - Figure 2.8. Similarly, the learning rate for punished trials was found to be lower for AN-R patients compared to the HC group (p = 0.0274, Cohen's d = 1.144) in the domain-specific block (Figure 2.9). However, the same difference did not emerge between the AN-R patients and the at-risk group (p = 0.0732). No credible difference was observed in the two learning rates for rewarded (p = 0.4325) and punished trials (p = 0.2327) between the AN-R group and the HC group, as well as between the AN-R group and the RI group (p = 0.3232; p = 0.2249), for choices made in the neutral block (Figure 2.10; Figure 2.11). Individuals with AN-R displayed a higher decision threshold (a) for food-related choices compared to both the HC group (p = 0.0068, Cohen's d = -0.622) and the RI group (p = 0.0118, Cohen's d = -0.454), while no credible group differences were found for neutral choices (Figure 2.12; Figure 2.13).

Lastly, no credible differences were found, both for within-group and between-group comparisons, concerning the parameters of cognitive processing speed (v) and motor response time (t).

2.5.4 Comorbidities

To investigate the influence of comorbidity and medication status on cognitive flexibility, a comparative analysis was conducted using Model M7 between AN-R participants diagnosed with comorbidities (comprising 45% of the sample) and those without any comorbid conditions. The comorbid group included patients with Generalized Anxiety Disorder (60%), Obsessive-Compulsive Disorder (30%), Social Phobia (5%), and Panic Disorder (5%). It is important to note that all participants receiving medication were also those with comorbid conditions, leading to an overlap between these two variables. Consequently, the analysis focused primarily on the presence of comorbidities rather than

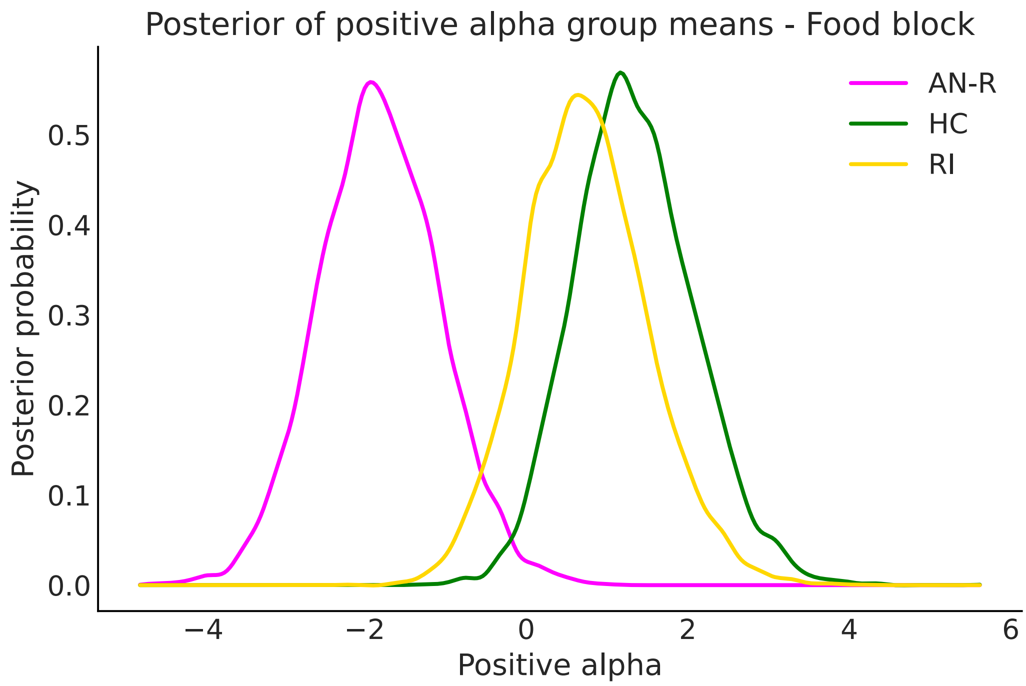

Figure 2.8: Posterior distributions comparison of the parameter α^{+} estimated separately for the three groups (AN-R, HC, RI) in the domain-specific block. The pink curve represents the posterior distribution of the α^{+} parameter for the AN-R group; the yellow curve represents the posterior distribution of the α^{+} parameter for the RI group; the green curve represents the posterior distribution of the α^{+} parameter for the HC group.

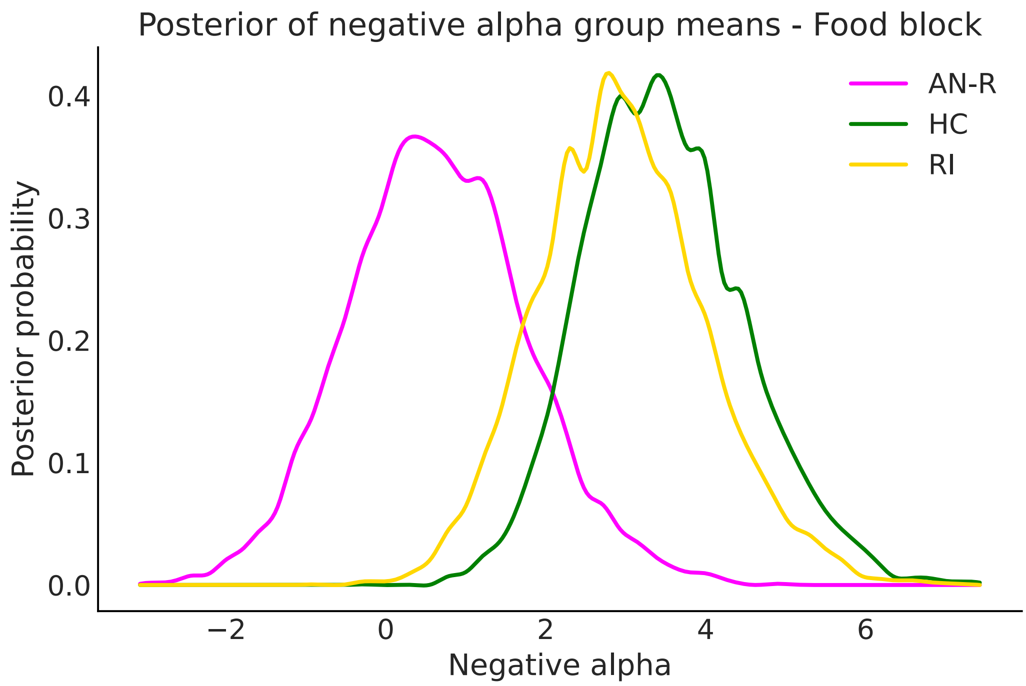

Figure 2.9: Posterior distributions comparison of the parameter α^{-} estimated separately for the three groups (AN-R, HC, RI) in the domain-specific block. The pink curve represents the posterior distribution of the α^{-} parameter for the AN-R group; the yellow curve represents the posterior distribution of the α^{-} parameter for the RI group; the green curve represents the posterior distribution of the α^{-} parameter for the HC group.

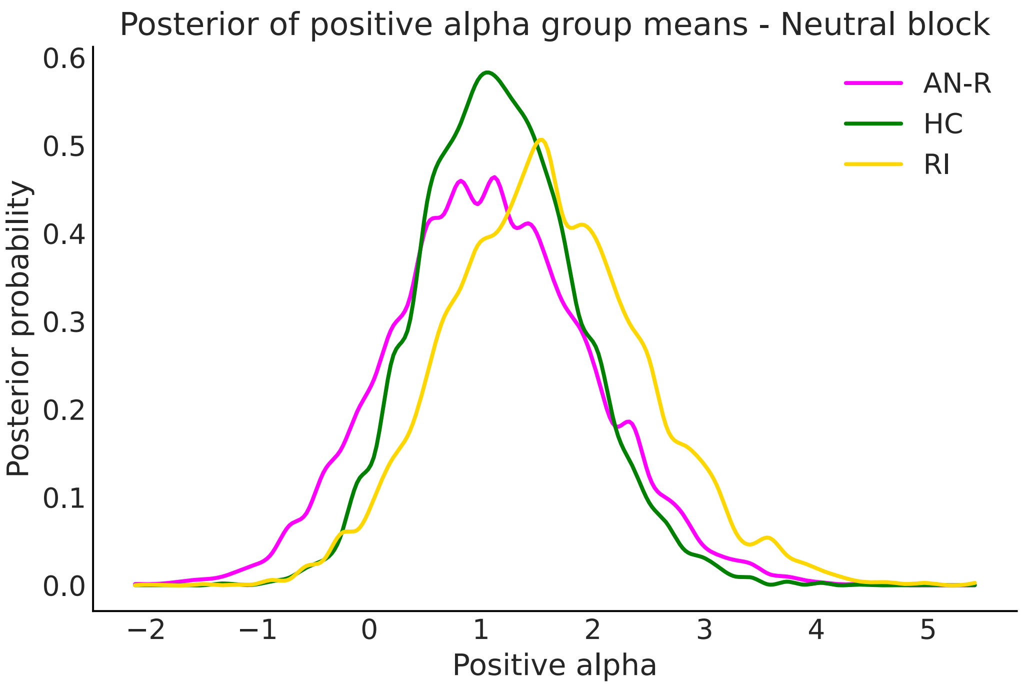

Figure 2.10: Posterior distributions comparison of the parameter α^+ estimated separately for the three groups (AN-R, HC, RI) in the neutral block. The pink curve represents the posterior distribution of the α^+ parameter for the AN-R group; the yellow curve represents the posterior distribution of the α^+ parameter for the RI group; the green curve represents the posterior distribution of the α^+ parameter for the HC group.

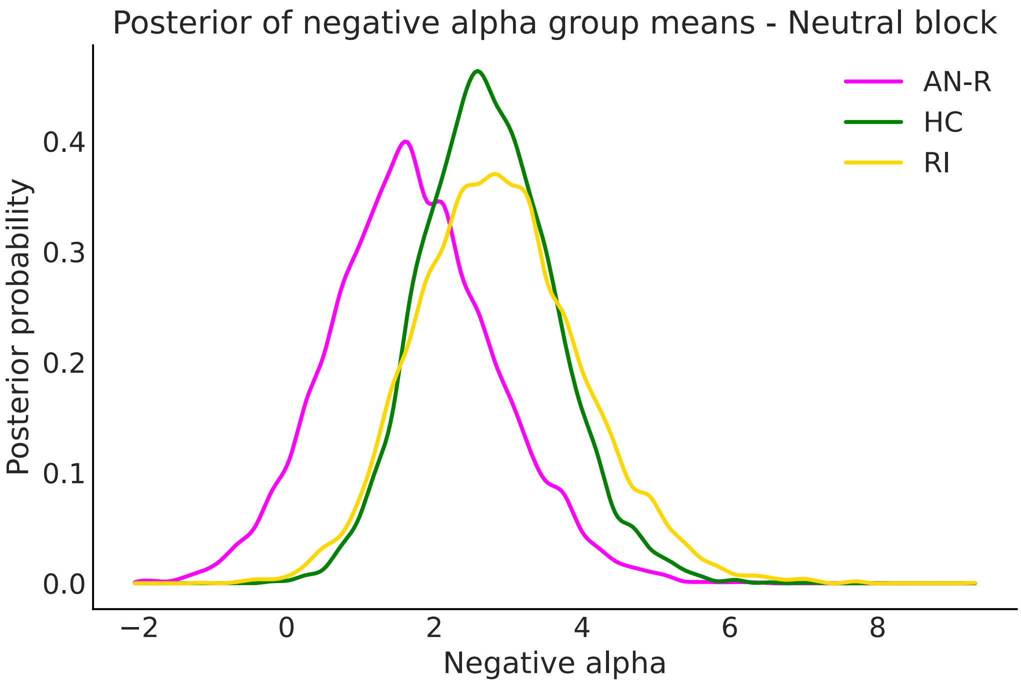

Figure 2.11: Posterior distributions comparison of the parameter α^- estimated separately for the three groups (AN-R, HC, RI) in the neutral block. The pink curve represents the posterior distribution of the α^- parameter for the AN-R group; the yellow curve represents the posterior distribution of the α^- parameter for the RI group; the green curve represents the posterior distribution of the α^- parameter for the HC group.

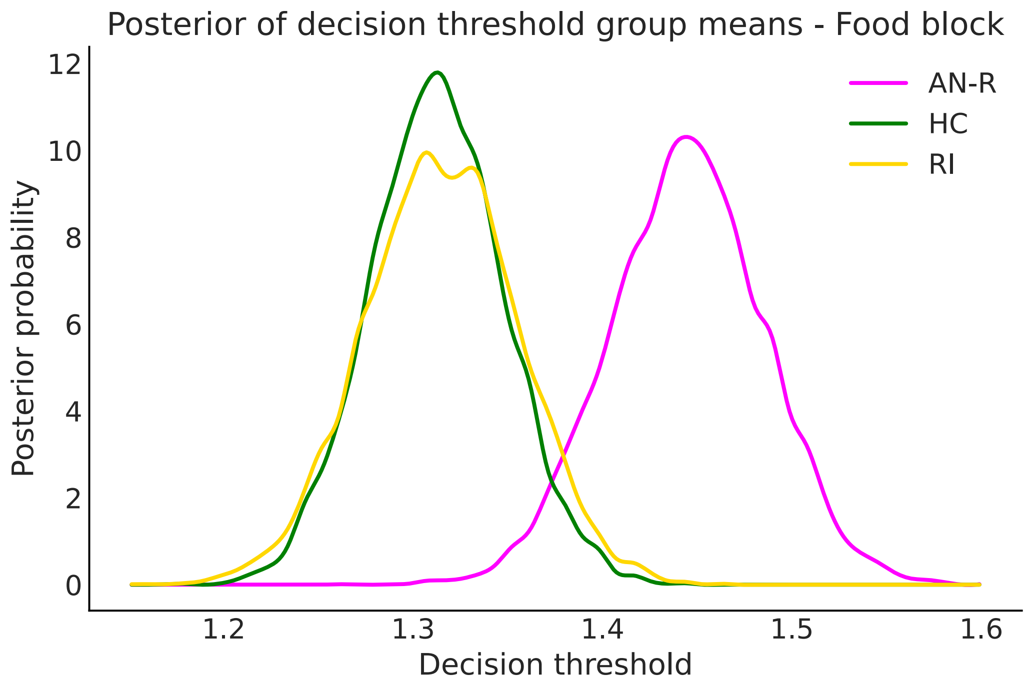

Figure 2.12: Posterior distributions comparison of the parameter a estimated separately for the three groups (AN-R, HC, RI) in the domain-specific block. The pink curve represents the posterior distribution of the a parameter for the AN-R group; the yellow curve represents the posterior distribution of the a parameter for the RI group; the green curve represents the posterior distribution of the a parameter for the HC group.

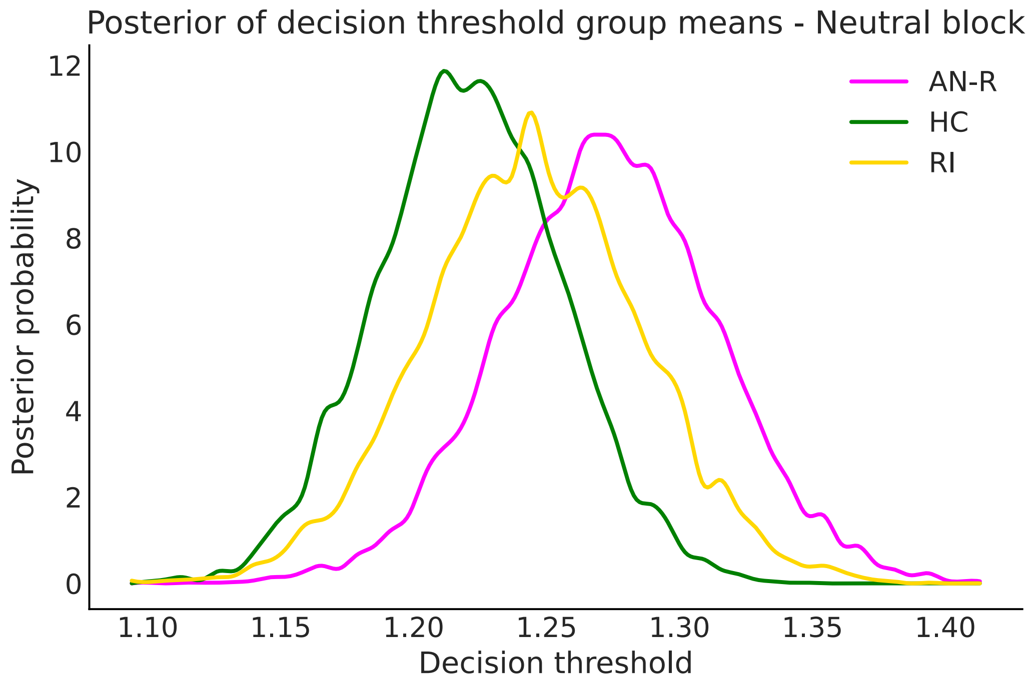

Figure 2.13: Posterior distributions comparison of the parameter a estimated separately for the three groups (AN-R, HC, RI) in the neutral block. The pink curve represents the posterior distribution of the a parameter for the AN-R group; the yellow curve represents the posterior distribution of the a parameter for the RI group; the green curve represents the posterior distribution of the a parameter for the HC group.

medication status. The correlation between comorbidity and medication was high (r = 0.78).

From the comparison of the estimated parameters, no credible differences were observed between participants with and without comorbidities. Specifically, considering the results from the ED-salient condition, the differences in the parameters were as follows: $\Delta\alpha^-$ = 2.614, 95% CI [-3.173, 8.364]; $\Delta\alpha^+$ = -0.635, 95% CI [-4.301, 2.449]; Δa = -0.034, 95% CI [-0.188, 0.124]; Δv = 0.230, 95% CI [-1.203, 1.586]; Δt = 0.002, 95% CI [-0.050, 0.055].

Similarly, for the neutral condition, the differences in the parameters were as follows: $\Delta\alpha^-$ = -0.768, 95% CI [-6.570, 4.401]; $\Delta\alpha^+$ = -1.739, 95% CI [-6.184, 1.654]; Δa = -0.126, 95% CI [-0.281, 0.025]; Δv = 0.744, 95% CI [-0.453, 1.886]; Δt = -0.003, 95% CI [-0.057, 0.052].

2.5.5 Correlation Analyses

Correlation analyses were conducted to explore the relationship between the key parameters of the HDDMrl, specifically α^- and α^+, and clinical measures directly associated with eating disorders, including EAT-26, BSQ-14, and MPS. These analyses included both patient and control groups to identify specific aspects within the domain of eating disorders related to reversal learning ability. The parameters α^- and α^+ were correlated with self-reported measures in scenarios involving both Food and Neutral choices, revealing notable differences for individuals with higher levels of eating dysfunctions. To elucidate these relationships, two distinct correlation matrices were generated: one for food-related choices and another for food-unrelated choices, underscoring the different correlations of eating disorder symptoms and learning processes in varied contexts.

In food choice scenarios, α^- showed a negative correlation with the total score of the EAT-26 and its subscales. This finding indicates that higher levels of disordered eating behaviors are associated with reduced learning from negative outcomes. Similarly, the positive learning rate (α^+) was negatively correlated with these measures. This suggests that individuals with higher EAT-26 total scores, particularly in the Oral Control subscale, tend to have a lower learning rate in scenarios where rewards are involved.

However, in neutral choice scenarios, these learning rates (α^- and α^+) exhibited weaker or negligible correlations with self-reported measures. This suggests that the strength of the correlation between psychological factors and learning processes is context-dependent, varying between food-related and neutral scenarios.

Finally, the strongest correlations with α^- and α^+ are related to dysfunctional eating habits, based on EAT-26 scores, rather than to body image concerns (BSQ-14) or perfectionism (MPS), especially in ED-salient scenarios. Detailed results of these analyses are available in the Appendix A.

2.6 Brief Discussion

The aim of the present study was to enhance the comprehension of cognitive flexibility deficits in AN, with a specific focus on determining whether these deficits are universal (domain-general) or context-specific (domain-specific). Cognitive flexibility plays a pivotal role in regulating reinforcement learning processes, which are vital for the adoption of appropriate and effective behaviors. Hence, understanding the mechanisms governing this ability is critical for optimal environmental adaptation (Walsh, 2013). Although rigidity is frequently observed in AN patients, it remains unclear whether this rigidity is predominantly influenced by ED-salient stimuli or is a broader response to various environmental cues. To shed light on this, we compared the performances of AN-R patients, healthy controls, and a group at risk of developing eating disorders—all matched for age, gender, and educational level—on two versions of a PRL task, one neutral and the other ED-salient.

Empirical Findings of the Current Investigation

Results revealed a context-dependent cognitive flexibility deficit in patients with AN-R compared to both the control group and the at-risk group. When comparing the performance of AN-R participants between the neutral and domain-specific blocks of the PRL task, a greater rigidity emerged in adopting a more advantageous decision-making strategy in the domain-specific block upon receiving positive feedback. However, this rigidity was not observed in the neutral block, where patients demonstrated greater decision-making flexibility, nor in response to punishments in both blocks. This performance difference between the two blocks was not found for either the control or the at-risk groups (*H1*). Moreover, comparing performances in the two blocks among the three groups, AN-R individuals exhibited lower learning rates in the block with ED-salient stimuli, both for rewarded choices and punished choices, compared to the HC group (*H2*). Similarly, AN-R individuals had a lower learning rate in rewarded trials compared to the at-risk individuals in the domain-specific block. However, no difference between the two groups emerged in the punished trials learning rate. Lastly, no differences in learning rates were observed between the three groups for choices unrelated to the disorder.

These results are further corroborated by the correlation analysis, which detected a strong negative correlation between dysfunctional eating habits and learning rates in food-related scenarios, but not in neutral scenarios. This observation supports the domain-specific nature of cognitive inflexibility.

Another important finding concerns the remaining parameters of the HDDMrl model. Particularly, the AN-R group showed a higher decision threshold (parameter a) compared to the HC and RI groups, but this difference only manifested in the context of food-related choices (*H3*). This suggests that AN-R individuals adopt a more cautious and conservative decision-making approach solely when making choices salient for individuals with ED (see Caudek et al., 2021; Schiff, Testa, Rusconi, Angeli, & Mapelli, 2021). However, when comparing performance between the two blocks within the same subject group, even healthy controls exhibited a more conservative strategy in the context of food-related choices compared to the neutral block. This finding is consistent with previous research, emphasizing that attention biases toward food-related stimuli are not exclusive to clinical populations but are also evident in the general population (Jonker, Glashouwer, Hoekzema, Ostafin, & Jong, 2019). As illustrated by Loeber et al. (2012), even non-clinical groups display significant attention shifts toward appetitive food indicators, underscoring the pervasive presence of this tendency.

Additionally, the data suggest that the observed results are unlikely to depend on the presence of a comorbid disorder alongside AN-R or medication use. This inference is drawn from the analysis undertaken to explore the interplay between model parameters and the incidence of comorbidities and medication use within the patient group (see the previous section on **Comorbidities** analysis).

Insights from the Results

The PRL task is considered especially suitable for investigating cognitive flexibility and reinforcement learning processes in laboratory settings (Fradkin et al., 2018). The task simulates an unpredictable and constantly evolving environment, repeatedly requiring participants to re-evaluate their actions based on received feedback. Indeed, performance in the PRL task is closely associated with two processes implying certain flexibility. Firstly, learning the stimulus-response association to facilitate adaptive decision-making. Secondly, promptly update expectations about the stimulus value and, accordingly, flexibly change the decision strategy where needed. The findings highlight that individuals with AN-R might struggle with both these processes. However, it is crucial to note that such difficulties emerge only in contexts associated with eating disorder symptoms and do not manifest during decision-making regarding neutral choices. This suggests that the learning deficit observed in individuals with AN-R is limited to ED-salient contexts and does not indicate a generalized impairment in reinforcement learning. In this particular scenario, the importance of context and the characteristics of the stimuli emerge as critical determinants in modulating the inherent mechanisms of reinforcement learning. This is surprising since, in the context of the PRL task, the specific nature of the presented stimulus should be irrelevant. Yet, for AN-R patients, the nature of the stimulus plays a pivotal role, challenging the cognitive flexibility mechanism, which is perfectly functioning in other contexts.

Finally, the results show a differential processing between rewards and punishments.

When comparing performances of AN-R patients across the two blocks, a slower adaptation is evident in the domain-specific block compared to the neutral block. However, this is only the case for choices where participants received a reward. No differences between blocks emerged for trials where participants were punished.

This finding aligns with prior studies suggesting that AN-R patients appear less sensitive to rewards, conferring them limited significance, but are hypersensitive to punishments, perceiving them intolerable (Bernardoni et al., 2021). Hence, it could be hypothesized that patients are highly motivated to avoid punishment, and may exert greater cognitive effort when faced with the possibility of punitive feedback, compensating for their performance in areas where they experience more challenges. On the other hand, it is plausible that rewards are perceived as minimally gratifying, especially when associated with contexts seen as threatening. In situations interpreted as threatening and punitive, individuals might resort to established, repetitive strategies providing a sense of control, even if not necessarily optimal for that specific context, neglecting information related to positive feedback (Haynos et al., 2020).

The mechanism described above, could represent a maintenance factor of the disorder (Steding et al., 2019; Walsh, 2013). In daily life contexts perceived as unpleasant, AN-R patients might employ the same behavioral strategy even when it becomes dangerous for health (*e.g.*, restrictive diets), misinterpreting environmental feedback that communicates the need for change (Walsh, 2013).

However, the extent to which challenges related to adaptability to a constantly changing environment could serve as early risk factors for preventive measures remains to be clarified. Notably, the at-risk group in the present study exhibited no discernible difficulties with the PRL task, mirroring the behavioral patterns observed in healthy controls.

In summary, the study suggests the presence of a domain-specific cognitive flexibility deficit in AN-R patients. This deficit seems to contribute to the perpetuation of the disorder rather than being a causal factor, as indicated by the results observed in the at-risk group. Recognizing and addressing this deficit is crucial in therapeutic interventions for AN-R.

Clinical Implications

The results from the current study may have considerable clinical implications. Nevertheless, it is essential to emphasize their preliminary nature. Rigorous replication is crucial, especially considering the potential ramifications these findings might have on the development of targeted therapeutic interventions in future research. Should convergent evidence emerge supporting a domain-specific deficit in cognitive flexibility among AN patients, a reassessment of existing therapeutic intervention strategies would be imperative.

The inability to optimally modify behaviors in contexts most relevant to AN-R patients may hinder therapeutic strategies. These strategies predominantly focus on supporting healthy eating habits and promoting psycho-physical well-being (Fairburn et al., 2003). Given this challenge, Cognitive Remediation Therapy (CRT, Tchanturia et al., 2010) has been introduced as an adjunct to conventional treatments, with the objective of enhancing cognitive flexibility in AN-R patients. CRT encompasses cognitive exercises and behavioral strategies designed to improve central coherence, mitigate cognitive and behavioral rigidity, and deepen comprehension of the thinking style (Tchanturia et al., 2010). A distinctive element of CRT is its intentional omission of ED-salient themes, choosing instead to use neutral stimuli during cognitive and behavioral tasks. This approach aims to strengthen the therapeutic alliance and reduce drop-out rates, particularly in AN-R patients. Yet, the effectiveness of CRT in enhancing central coherence, cognitive flexibility, or symptoms related to eating disorders has been recently questioned (Hagan et al., 2020; Tchanturia, Giombini, Leppanen, & Kinnaird, 2017). In line with these recent insights, Trapp, Heid, Röder, Wimmer, & Hajak (2022) recommends modifications to address the practical challenges associated with the application of CRT. Specifically, drawing from Beck's cognitive depression theory (Beck & Alford, 2009), the authors advocate shifting from solely using neutral stimuli to incorporating emotionally-charged stimuli, thereby inducing "warm" cognitive responses, within the cognitive training framework. Consequently, in the context of AN-R, this implies that the incorporation of ED-salient stimuli might enhance cognitive flexibility within this therapeutic approach.

It is pivotal to reiterate the preliminary nature of the study. If future research substantiates the domain-specific hypothesis of the deficit, a shift in therapeutic intervention towards the aforementioned direction will be necessitated.

Limitations and Future Directions

This study is not without limitations.

1) The study focused on AN-R outpatients from a secondary level eating disorders care center, suggesting that the severity of their symptoms likely ranged from medium to low. It is crucial to underscore that the present findings apply specifically to this group, and evaluating patients with different characteristics might yield varying results. For instance, AN-R patients with severe symptoms might exhibit a generalized impairment in reinforcement learning across various decision-making contexts, implying that as the illness progresses, the cognitive deficit could transition from being domain-specific to domain-general. This could imply a need for varied therapeutic approaches based on symptom severity. Exploring cognitive flexibility in diverse AN-R patient groups, such as those at advanced disease stages, is identified as a key area for future research (**Study 2**).

2) While this study involved AN-R patients, the *transdiagnostic* hypothesis suggests that cognitive flexibility deficit could span multiple diagnostic categories. Broadening the scope to include other eating disorders like AN Binge-Eating subtype, BN, and BED is crucial to further understand this phenomenon.

3) The precise influence of rewards and punishments on AN-R patients' reinforcement learning difficulties is still unclear. Although this study considered this aspect, the main focus was on manipulating contextual stimuli rather than feedback. The feedback used was symbolic and did not account for its subjective value to individuals. Future research could employ tangible rewards and punishments with universal significance for all participants.

4) This study investigates deficits in cognitive flexibility as potential underlying factors for both the maintenance and development of AN-R. Evidence suggests that impaired cognitive flexibility, discernible from deficits in reversal learning, may contribute to the persistence and therapeutic complexities of AN-R. However, its role as a direct risk factor in the onset of the disorder is less clear. In this study, the AN-R group exhibited a deficit in cognitive flexibility compared to HC in the context of ED-salient cues, suggesting its underlying role in maintaining the disorder. In contrast, the performance of the at-risk group in the PRL task was comparable to that of HC, suggesting that this cognitive deficit may not manifest in the early or premorbid stages of the disorder. This finding leads to the hypothesis that cognitive inflexibility likely emerges in the later stages of AN-R, rather than functioning as an initial risk factor. On the other hand, a larger-scale study by Caudek et al. (2021) found that at-risk individuals exhibited a domain-specific cognitive flexibility deficit, differing from the non-at-risk group. This inconsistency highlights the imperative for more expansive and varied research to elucidate the impact of cognitive flexibility on both the onset and maintenance of eating disorders.

5) Future research should ideally adopt a longitudinal design, diverging from the cross-sectional approach of the current study. Such longitudinal investigations, tracking cognitive flexibility from the early stages through to the more advanced phases of the disorder, would be instrumental in determining the role of cognitive flexibility as a potential underlying mechanism for both the maintenance and onset of eating disorders. Understanding the temporal dynamics of cognitive processes is crucial for a clearer interpretation of cognitive flexibility impact on these disorders.

Chapter 3

Study 2: Exploring Cognitive Flexibility in Severe Eating Disorders: A Comparison Between Anorexia Nervosa and Bulimia Nervosa

3.1 Introduction

Eating disorders (EDs) are multifaceted pathological conditions, with a vast spectrum of symptoms and clinical outcomes. Understanding their complexity, variations in symptom severity, and underlying cognitive processes can provide deeper insights into diagnosis and treatment. In this context, we primarily explore the influence of symptom severity on cognitive processes within EDs.

3.1.1 Complexity and Spectrum of Eating Disorders

Eating disorders manifest with a broad range of symptoms, leading to a diverse set of clinical presentations. Variations in symptom severity necessitate different treatment approaches. While milder cases may benefit from outpatient interventions, such as individual therapy, nutritional guidance and psychological assistance, severe cases often require medical attention, intensive psychological intervention, and even hospitalization. The mortality rate can be considerably high, even reaching up to 20% (Qian et al., 2022).

3.1.2 Assessing Severity of the Disorder

Several criteria are taken into account to determine the severity of the disorder:

(1) Body Mass Index (BMI), notably for assessing severity in AN patients, as significant underweight status can result in pronounced physiological and psychological consequences. A BMI of 17.5 or lower is classified as severe underweight, demanding medical attention (American Psychiatric Association, 2022; Fairburn et al., 2003);

(2) vital parameters, including blood parameters (electrolyte levels, proteins, hemoglobin, and blood sugar), blood pressure, heart rate, body temperature, bone density, and the overall condition of the skin, hair, and nails (Forney, Buchman-Schmitt, Keel, & Frank, 2016; Gallagher, Parker, Samavat, & Zelig, 2022; Mehler & Brown, 2015);

(3) psychiatric comorbidity serves as an additional indicator of severity. The concurrent presence of another psychiatric disorder, such as depression or anxiety, can further complicate an already unstable condition (American Psychiatric Association, 2022);

(4) aspects more closely related to the disorder itself by evaluating key ED dimensions (*e.g.*, food restriction, frequency of binge-eating and purging behaviors, body dissatisfaction, etc.), often measured using self-report tools or clinical interviews like the Eating Disorder Examination Questionnaire (EDE-Q, Fairburn & Beglin, 1994), the Structured Clinical Interview for Eating Disorders (SCID-5, American Psychiatric Association, 2013), and the Yale-Brown-Cornell Eating Disorder Severity Scale (YBC-EDS, Sunday, Halmi, & Einhorn, 1995).

(5) the degree of functional impairment and the impact on quality of life (American Psychiatric Association, 2022).

3.1.3 The Role of Cognitive Flexibility in Severe Anorexia Nersosa

Despite the emphasis on medical and psychological dimensions, the crucial role of underlying cognitive processes often remains neglected (Davis et al., 2020). However, as symptoms worsen, cognitive deficits may intensify, forming a vicious cycle that perpetuates the disorder (Walsh, 2013). This highlights the importance of considering cognitive deficits when assessing the severity of EDs, especially AN. Visu-Petra & Mărcuș (2019) suggested that cognitive inflexibility might serve as a persistent factor in psychopathology. In line with this perspective, increased symptoms could exacerbate cognitive rigidity, thus hindering therapeutic progress. An intriguing observation is that while milder AN cases exhibit domain-specific cognitive flexibility deficits (as outlined in **Study 1**), in more severe cases, this deficit might become domain-general.

Previous studies have investigated the link between chronic, severe eating disorders and cognitive inflexibility in a domain-general perspective. Tchanturia et al. (2012), when comparing performances on a classic version of the WCST task in groups of patients diagnosed with AN across different severity levels, found that patients with more severe symptoms tend to display greater cognitive rigidity. This rigidity potentially hinders changes in dysfunctional eating behaviors. In contrast, patients with milder symptoms seemingly exhibited enhanced adaptability and cognitive flexibility.

Similarly, it has been suggested that the consolidation of dysfunctional eating habits due to a reinforcement learning deficit is more accentuated in patients with AN showing a more severe clinical profile (Davis et al., 2020; Walsh, 2013). Using self-report measures, Davis et al. (2020) found that patients with AN, especially those with longer illness duration and heightened symptom severity reported more rigid eating-related habits and associated behaviors, which were difficult to modify. Over time, individuals with AN become insensitive to any environmental feedback, leading them to automatically relapse on previously learned behaviors, even when such behaviors are no longer adaptive. This occurs irrespective of the disastrous consequences on their psycho-physical health (Walsh, 2013).

In addition, symptom severity can also impact other cognitive processes, such as selective attention. In a study by Giel et al. (2011), it was observed that patients with severe AN-R displayed greater attentional biases, in response to stimuli associated with food and body weight, compared to less severe or remitted patients. The study suggests that symptom severity may alter information processing in this patient population. Further studies have explored the impact of eating symptom severity on inhibitory control (*i.e.*, defined as the ability to suppress or override automatic responses when they are inappropriate in a given context). Again, illness duration and a worsening of eating symptoms, coupled with general physical debilitation, lead to a greater impairment in inhibitory control processes. Specifically, in the case of AN, this impairment manifests as an excessive reliance on inhibitory controls, potentially hindering adaptive behavioral responses (Miranda-Olivos et al., 2021).

Given these findings, it is crucial to examine cognitive deficits across varying symptom severity levels. This can provide deeper insights into whether these deficits, in the context of the persistence and progression of the disorder, are domain-specific or domain-general.

3.1.4 Cognitive Inflexibility as Transdiagnostic Factor: A Comparison Between AN and BN

Cognitive dysfunctions can also differ significantly depending on the specific diagnostic category under consideration. For instance, AN and BN are two of the most extensively studied and recognized eating disorders in scientific literature. Although AN and BN share certain features, such as an excessive concern about weight and body shape (Fairburn et al., 2003), it is important to acknowledge the significant differences in their cognitive profiles, which have been underscored by prior research. These differences reinforce the notion that AN and BN, despite having some common core features, are fundamentally distinct disorders (Curzio et al., 2018). For instance, it is well-documented that AN and BN patients differ in their inhibitory control ability, which refers to the capacity to suppress impulsive and undesired responses (Bartholdy et al., 2016; Diamond, 2013). Numerous studies, including Bartholdy et al. (2016), have explored this distinction.

Traditionally, BN is linked with impulsive behaviors, such as binge episodes and their associated compensatory actions. In contrast, AN patients exhibit stronger self-regulation and superior inhibitory control.

On the other hand, in accordance with the *transdiagnostic* perspective of cognitive inflexibility (Morris & Mansell, 2018), both AN and BN are believed to demonstrate this specific cognitive dysfunction, albeit with some variations in its manifestation, attributable to their inherent differences. Nonetheless, studies investigating cognitive flexibility in AN and BN have reached inconclusive results.

Roberts et al. (2007) observed that both BN and AN patients demonstrated poor set-shifting abilities compared to healthy controls across multiple cognitive tasks. However, contrasting findings were presented by Darcy et al. (2012), who identified no differences in cognitive flexibility, as evaluated by performance on the WCST, between BN patients and healthy controls. Similar outcomes were reported by both Galimberti et al. (2012) using the Intra/Extra-dimensional set-shifting task, and Perpiñá et al. (2017) using the WCST, in the context of BN. Despite these results, Hirst et al. (2017) observed deficits in cognitive flexibility in both AN and BN across various studies. They postulated that this deficit might be even more pronounced in BN patients than in those with AN.

Finally, recent studies have underscored relevant differences in reinforcement learning between AN and BN patients. These studies observed that individuals with AN possess pronounced cognitive control, leading them to rigidly adhere to reinforcement learning patterns in their efforts to exert strict control over food and weight (Wonderlich, Bershad, & Steinglass, 2021). In severe cases, this behavior can evolve into deeply entrenched habitual patterns, making environmental feedback irrelevant (Davis et al., 2020). Conversely, BN patients tend to exhibit weaker cognitive control. They are deeply influenced by a heightened sensitivity to positive feedback, such as the euphoria following a binge, and are more susceptible to negative feedback, experiencing intensified feelings of guilt and shame afterward (Chan et al., 2014; Danner, Evers, Stok, Elburg, & Ridder, 2012b; Hagan & Forbush, 2021; Lee, Namkoong, & Jung, 2017). Thus, although through different mechanisms, both groups demonstrate alterations in their reinforcement learning abilities.

This brief literature review focusing on cognitive flexibility deficits in AN and BN underscores contradictory findings of previous research. Given these inconsistencies, there is a clear imperative to delve deeper into this area of study, aiming to elucidate the *transdiagnostic* nature of cognitive inflexibility, but also its variations and the broader implications for each diagnostic category.

The discrepancies observed in previous research regarding cognitive flexibility in AN and BN may be attributable to the insufficient consideration of disorder-specific characteristics, particularly symptom severity. There is a growing hypothesis that as symptom severity escalates, cognitive deficits may manifest differently for both AN and BN. This notion is supported by findings from Mora-Maltas et al. (2023), who observed that patients with severe AN exhibited more pronounced flexibility deficits in task performance compared to those with a shorter duration of the disorder. Similarly, in BN, longer illness duration correlated with greater impairments in cognitive flexibility.

To elucidate these patterns more clearly, it is imperative to focus on more homogeneous groups concerning symptom severity. This approach will enable a more precise delineation of how cognitive flexibility varies within each subgroup. Particularly in severe cases, cognitive flexibility deficits are expected to be more pronounced, contrasting with milder manifestations observed in less severe cases. Such targeted research could yield enhanced clarity into the *transdiagnostic* nature of cognitive flexibility (Grau, Magallón-Neri, Faus, & Feixas, 2019; Mora-Maltas et al., 2023).

Moreover, a detailed examination of individuals with severe AN and BN is crucial for identifying unique cognitive traits that differentiate the two disorders, particularly in the realm of cognitive flexibility. This investigation is key to determining whether these cognitive characteristics remain distinct even with increased severity of the disorders. This understanding is important for substantiating the hypothesis that AN and BN, while sharing commonalities as eating disorders (Fairburn et al., 2003), are fundamentally separate entities with unique cognitive profiles. Such findings would not only reinforce the distinct nature of these disorders but also highlight the importance of tailored approaches in treatment and research, acknowledging the specific cognitive nuances of each condition.

Therefore, if a deficit in cognitive flexibility were to emerge in both severe cases of AN and BN, albeit with slight differences, it would underscore two important aspects. To begin with, the occurrence of cognitive flexibility deficits in both disorders suggests that cognitive inflexibility might be a *transdiagnostic* factor prevalent across various forms of psychopathology. Next, the distinct ways in which this deficit presents in these two conditions would underscore the fundamental differences between AN and BN. This means that while they fall under the umbrella of eating disorders, each exhibits unique characteristics, requiring distinct approaches in treatment and understanding (Curzio et al., 2018). Therefore, the presence of cognitive flexibility deficits across AN and BN, despite their distinct nature, reinforces the idea that these deficits are a key component in both disorders.

In summary, there are two pivotal issues to address. Firstly, it is important to consider the impact of symptom severity on cognitive flexibility in AN patients. **Study 1** revealed that, among patients with milder symptoms, this deficit appears in a domain-specific manner and not in a domain-general way. This suggests that, for this specific group, the deficit is confined to contexts directly related to their symptoms. However, as the disorder advances, the cognitive deficits might become more pronounced. For patients with more severe symptoms, this inflexibility might become domain-general, suggesting a more widespread cognitive impairment. Secondly, in order to test the *transdiagnostic* nature of cognitive flexibility, an exhaustive comparison between BN and AN patients, especially those with severe symptoms, is crucial. This comparison primarily aims to examine their capacity for flexible adaptation to evolving environmental conditions. It is anticipated that this ability may be impaired in both groups, but the extent and nature of these impairments are likely to vary due to the unique characteristics inherent to AN and BN, which are separate and distinct disorders.

In the current study, performances on a cognitive flexibility task were evaluated among groups: individuals diagnosed with severe AN, those with severe BN, and a comparison between those with less severe AN and severe AN. A novel emphasis was placed on examining groups with notably pronounced symptom severity.

3.2 The Present Study

The primary goal of the current study is to investigate the impact of symptom severity on cognitive flexibility deficits within Anorexia Nervosa Restrictive type (AN-R) patients. Specifically, the study seeks to determine if the cognitive deficit becomes pervasive across all contexts (*i.e.*, domain-general) as the illness progresses. Additionally, it aims to identify any potential deficit in cognitive flexibility in patients with BN.

Based on prior research, it is plausible to anticipate that: (1) In severe cases, cognitive flexibility deficits might manifest universally across contexts, suggesting a domain-general rather than domain-specific impairment; (2) while both severe AN and BN patients will exhibit cognitive inflexibility, being a *trasdiagnostic* factor, unique differences might still emerge between the two groups.

To answer these questions, two methodological approaches were established:

Firstly, it was investigated whether inpatients with severe eating disorders display a domain-general cognitive flexibility deficit, in contrast to the domain-specific deficit observed in less severe cases (**Study 1**). The comparison involved a group of severe AN patients and another group of AN patients with milder symptom severity, from **Study 1**, using a neutral version of the PRL task. This investigation stems from the idea that as the disorder intensifies, an individual's capacity to adapt to changing environments diminishes (Davis et al., 2020).

Secondly, the investigation turned to the cognitive flexibility deficit in patients with severe BN. While AN is frequently associated with cognitive inflexibility, BN has been less thoroughly studied, leading to inconsistent results (Darcy et al., 2012; Hirst et al., 2017). In alignment with the *transdiagnostic* perspective proposed by Visu-Petra & Mărcuș (2019), it is hypothesized that patients with severe BN might exhibit a reinforcement learning deficit, although different from that observed in AN-R patients. A group of severe AN and BN patients was compared using the PRL task.

Performance deficits in patients are anticipated across both diagnostic groups, consistent with previous studies emphasizing cognitive decline correlating with symptom progression.

Specifically, the following hypotheses are posited: *H1*: Patients with severe AN-R will exhibit a reduced learning rate in the PRL task compared to both healthy controls (HC) and AN-R outpatients with milder symptoms. *H2*: Patients with severe BN will exhibit a reduced learning rate in the PRL task compared to HCs. *H3*: BN patients, when performing the PRL task, will exhibit greater impulsivity than both HCs and AN-R patients. *H4*: Patients with severe AN-R will experience extended delays in processing the task stimuli, manifesting in both cognitive and motor responses.

To test the hypotheses, a neutral PRL task, without ED-salient stimuli, was administered to several groups: inpatients with AN-R, outpatients with AN-R, those with BN, and healthy controls. The collected data were then analyzed using the *Hierarchical Reinforcement Learning Drift Diffusion Model* (HDDMrl, Pedersen & Frank, 2020). This sophisticated computational model enables a detailed examination of the underlying mechanisms in the reinforcement learning process during the PRL task.

3.3 Method

3.3.1 Participants

The sample consists of 34 inpatient diagnosed with AN-R (M_{age} = 23.44, SD_{age} = 10.99), 36 outpatients with AN-R from **Study 1** (see **Study 1** for details), 21 inpatient diagnosed with BN (M_{age} = 22.63, SD_{age} = 8.37), and 115 HC (M_{age} = 20.53, SD_{age} = 2.37). The four groups include only female participants. Inpatients of the present study were recruited from the Italian center for intensive treatment of eating disorders, Villa dei Pini (Florence, Italy). All recruited patients met the diagnostic criteria for an eating disorder according to the DSM-5 (American Psychiatric Association, 2013). Diagnostic evaluations were undertaken by psychiatric and psychological professionals within the mentioned clinical center. In addition to the primary diagnosis of ED, concurrent psychiatric comorbidities were also assessed through additional psychological evaluations. The comprehensive diagnostic assessment revealed that 15% of the recruited patients presented with comorbidities involving other psychiatric conditions. Moreover, the majority of them were taking medications. The same professionals assessed whether the patient group met the inclusion criteria through semi-structured clinical interviews. Exclusion criteria included the presence of neurological disorders, suicidal ideation, substance or alcohol addiction, and psychosis. The HC group was voluntarily recruited through social media or university announcements. Subsequently, HC participants underwent the Eating Attitudes Test-26 (EAT-26) screening questionnaire to assess potential inclinations towards developing eating disorders. Those who scored above the cut-off threshold (≥ 20) on the EAT-26 were considered at risk of developing eating disorders and were excluded from the research. Exclusion criteria included abnormal BMI values, current treatment or diagnosis of an ED, neurological disorders, suicidal ideation, substance or alcohol addiction, and psychosis. All participants displayed normative cognitive functioning, as determined by the Raven's Standard Progressive Matrices test (Raven, 2003). All participants were of Caucasian ethnicity, right-handed, proficient in reading/writing and understanding the Italian language, and were kept unaware of the purpose of the study.

3.3.2 Materials

Probabilistic Reversal Learning (PRL) Task

In the present study, a PRL task with neutral stimuli (*e.g.*, a lamp) was administered. The use of neutral stimuli was based on the rationale that while cognitive deficits in moderate to low severity AN-R patients are evident only when exposed to ED-salient stimuli (**Study 1**), those with pronounced symptom severity are expected to demonstrate such deficits with food-related stimuli. Yet, it remains to be determined whether these impairments extend to all contexts, including neutral ones, as the disorder progresses. In the context of this study, the focus was on the domain-general nature of CI.

Stimuli were selected from the IAPS database (Lang & Bradley, 2007): five images of flowers (IAPS #5000, 5001, 5020, 5030, 5202) and five images of objects (IAPS #7010, 7020, 7034, 7056, 7170). The PRL block involves presenting two images per trial (randomly taken from the two aforementioned categories) for 160 trials. The 160 trials are divided into 4 epochs of 40 trials each. In each epoch, one of the two stimuli is rewarded 70% of the time, while the other 30%. Upon changing epochs, the stimulus-reward contingency is reversed. The reward is symbolized by an image of a symbolic euro coin, whereas the punishment is represented by an image of a crossed-out symbolic euro coin. The experiment was programmed and implemented through the online Psytoolkit platform (https://www.psytoolkit.org).

3.3.3 Procedure

The data collection involved three phases. In the first phase, the inclusion and exclusion criteria were examined to select the healthy control group. Those who voluntarily chose to participate in the study as control subjects were asked to provide their weight and height to calculate the BMI, and to complete the EAT-26 screening test. Participants scoring above the cut-off (≥ 20) were excluded from the final sample. Subsequently, a further evaluation was conducted to exclude the presence of neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis. Out of the initial 122 HC participants, 7 were excluded based on scores above the cutoff for the EAT-26 test.

Patients underwent a similar assessment by mental health professionals at the recruiting clinical site. In the second phase, all participants completed the Raven's Standard Progressive Matrices test (Raven, 2003) to assess their overall cognitive functioning. Finally, in the third phase, participants performed the Probabilistic Reversal Learning task.

Each phase took roughly 30 minutes on average. Participation in the study was voluntary, and no incentives of any kind were provided to the participants. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Research Ethics Committee of the University of Florence (Prot. n. 0178082). All participants provided informed consent, received privacy policy, and agreed to participate in the research.

3.4 Data Analysis

Data analysis was performed using the software Python (https://www.python.org) and R version 4.2.1 (https://cran.r-project.org/index.html).

As in **Study 1**, the PRL data was analyzed by estimating the 6 parameters of the computational model *Hierarchical Reinforcement Learning Drift Diffusion Model* (HDDMrl, Pedersen & Frank, 2020). However, in this study, parameters were estimated for each group involved but not for different stimuli, having solely a neutral version of the PRL task. Therefore, the model selected here, is simpler than the model employed in **Study 1**.

Between-groups comparison of estimated parameters. The parameters, estimated distinctly for each group, were subsequently compared. Four different group comparisons were undertaken: (1) a comparison between the AN-R inpatient group and the control group, (2) a comparison between the BN inpatient group and the control group, (3) a comparison between the AN-R inpatient group and the BN inpatient group, and (4) a comparison between the AN-R inpatient group and the AN-R outpatient group with mild to moderate severity.

3.5 Results

3.5.1 Quality Control

To exclude participants who may have performed the PRL task randomly, a quality control was conducted on the PRL data for all three groups. Participants who completed the task without following a criterion that allowed them to make choices above chance level were excluded from further analyses. A total of N = 160 participants satisfied the quality control criteria (reaction times between $\geq$ 150 ms and $\leq$ 2500 ms; accuracy below 40%; absent responses), comprising N = 30 individuals diagnosed with AN-R, N = 20 with BN, and N = 110 HCs. Further analyses were performed on this subset of participants.

3.5.2 Model Selection

In this case (as in **Study 1**), a comparison of HDDMrl models was necessary to select the model that represents the best trade-off between model fit to the data and model complexity. For the present study, the model exhibiting the lowest DIC was identified as the most complex model (M6). This model posits two distinct learning rates for rewards and punishments. It estimate all parameters separately for the three evaluated groups (AN-R, BN, HC). Again, the parameter z, pertaining to response bias, was not useful (for details see Appendix B).

3.5.3 Results from the HDDMrl Model

The HDDMrl Model (M6) was estimated using 15,000 iterations, with a *burn-in* of 5,000 iterations. The $\hat{R}$ values for all estimated parameters were below 1.1, indicating that the model achieved good convergence.

To explore a potential deficit in cognitive flexibility among patients with Eating Disorders, the posterior estimates of the HDDMrl parameters were compared across the specified groups. Specifically, parameter estimates for the AN-R group were contrasted with those of the controls, just as estimates for the BN group were set against the controls. A further comparison was made between the AN-R and the bulimic patient groups to discern differences in task performance across these diagnostic categories. Moreover, by contrasting the severe AN-R group with the AN-R group of milder symptom severity, the intent was to directly evaluate the impact of symptom severity on cognitive flexibility.

Between-groups comparison of estimated parameters: AN-R and Healthy Controls

Results show that the AN-R patients exhibited a reduced learning rate (α) both in punished trials (p = 0.0371, Cohen's d = 1.222) and in rewarded trials (p = 0.0227, Cohen's d = 0.966) compared to healthy controls (Figure 3.1; Figure 3.2). Individuals with AN-R displayed a higher decision threshold (p = 0.0, Cohen's d = -1.353), increased slowness in cognitive processing (v) of the two stimuli (p = 0.0078, Cohen's d = 1.088),

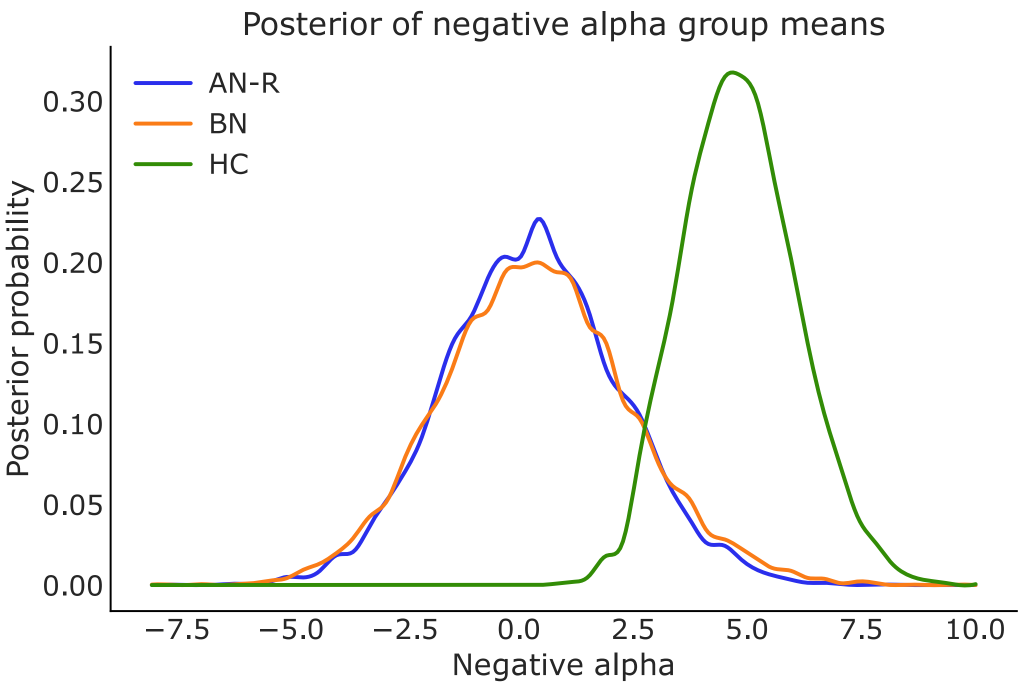

Figure 3.1: Posterior distributions comparison of the parameter α^- estimated separately for the three groups (AN-R, BN, HC). The blue curve represents the posterior distribution of the α^- parameter for the AN-R group; the orange curve represents the posterior distribution of the α^- parameter for the BN group; the green curve represents the posterior distribution of the α^- parameter for the HC group.

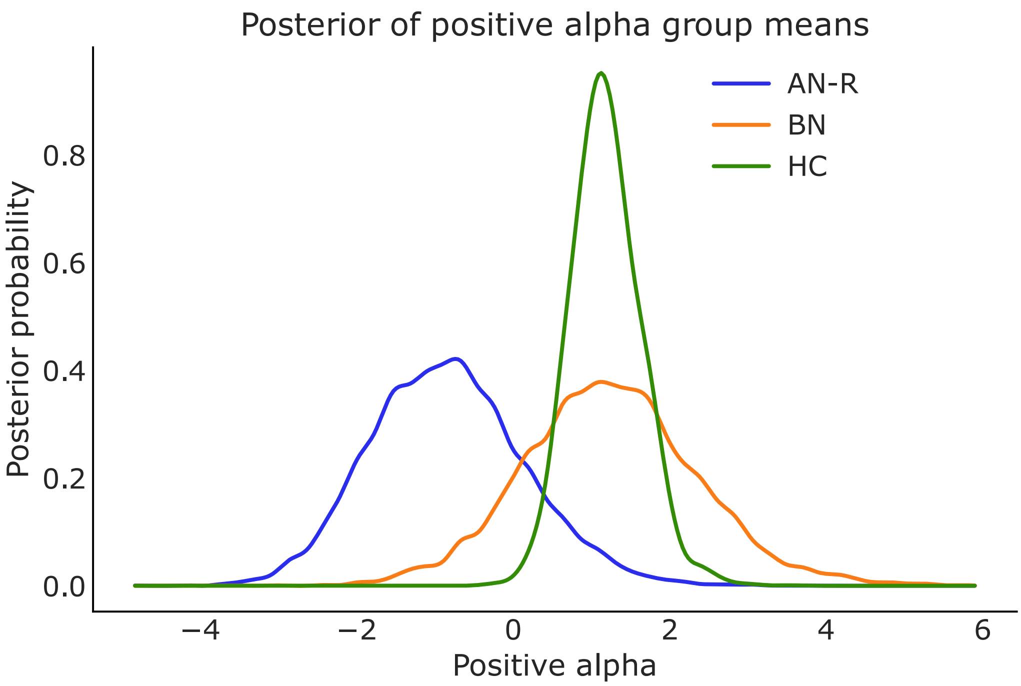

Figure 3.2: Posterior distributions comparison of the parameter α^+ estimated separately for the three groups (AN-R, BN, HC). The blue curve represents the posterior distribution of the α^+ parameter for the AN-R group; the orange curve represents the posterior distribution of the α^+ parameter for the BN group; the green curve represents the posterior distribution of the α^+ parameter for the HC group.

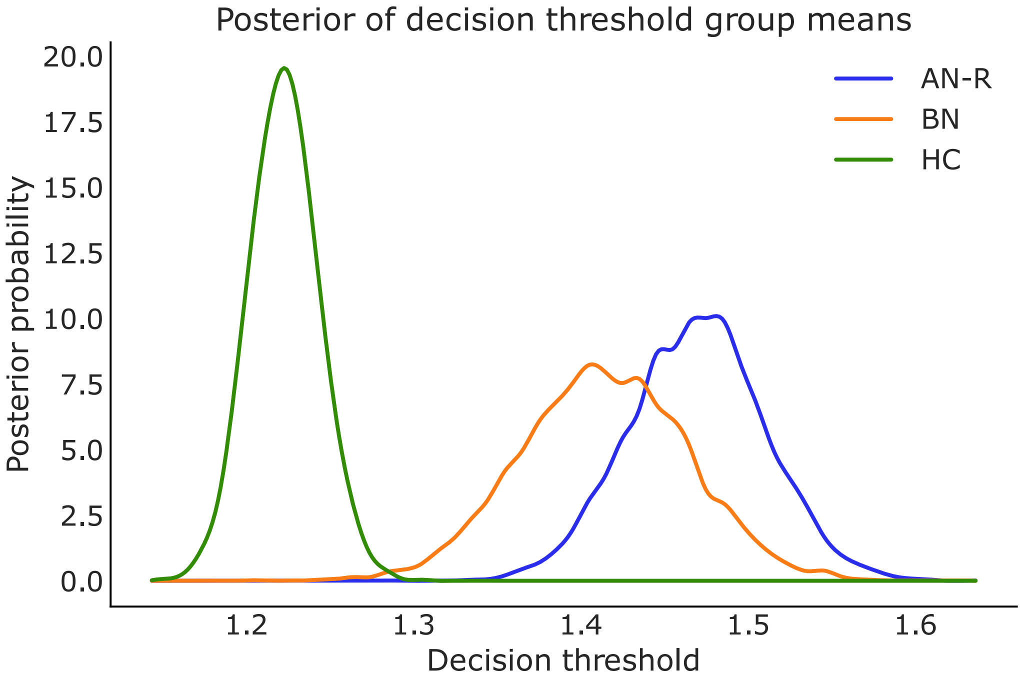

Figure 3.3: Posterior distributions comparison of the parameter a estimated separately for the three groups (AN-R, BN, HC). The blue curve represents the posterior distribution of the parameter a for the AN-R group; the orange curve represents the posterior distribution of the parameter a for the BN group; the green curve represents the posterior distribution of the parameter a for the HC group.

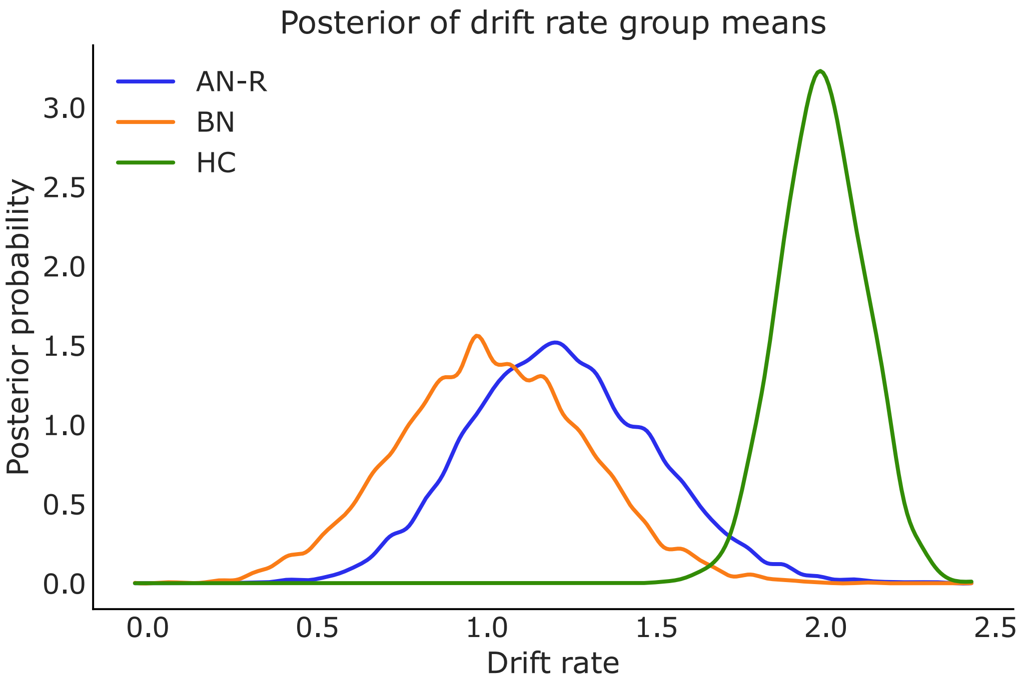

Figure 3.4: Posterior distributions comparison of the parameter v estimated separately for the three groups (AN-R, BN, HC). The blue curve represents the posterior distribution of the parameter v for the AN-R group; the orange curve represents the posterior distribution of the parameter v for the BN group; the green curve represents the posterior distribution of the parameter v for the HC group.

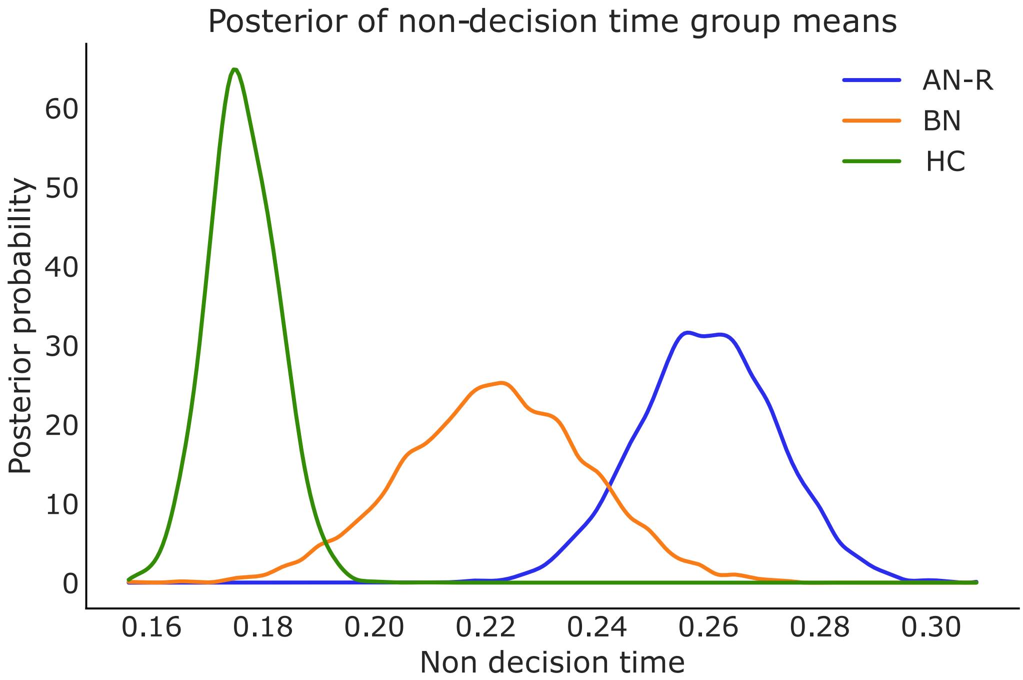

Figure 3.5: Posterior distributions comparison of the parameter t estimated separately for the three groups (AN-R, BN, HC). The blue curve represents the posterior distribution of the parameter t for the AN-R group; the orange curve represents the posterior distribution of the parameter t for the BN group; the green curve represents the posterior distribution of the parameter t for the HC group.

and a greater delay in the perception of the two stimuli followed by the consequent motor decision (t) (p = 0.0, Cohen's d = -1.685) in comparison to healthy controls (Table 3.1). Differences in the parameters posterior distribution between AN-R and HC are illustrated in the following plots: Figure 3.3 (parameter a), 3.4 (parameter v), and 3.5 (parameter t).

Table 3.1: Posterior Estimates of the HDDMrl Model Parameters by Group (AN-R, HC). The learning rates (α) are displayed on a logit scale. The probability (p) describes the Bayesian test comparing the posterior estimate of the parameter for the AN-R group to that of the HC group. Standard deviations are provided in parentheses. Cohen's d represents the effect size measure for the group comparison.

Parameter	AN-R Mean (SD)	HC Mean (SD)	p	Cohen's d
a	1.469(0.203)	1.219(0.179)	0.0000	-1.353
v	1.218(0.663)	5.053(0.704)	0.0078	1.088
t	0.188(0.081)	1.183(0.051)	0.0000	-1.685
α^-	0.353(5.407)	0.168(4.709)	0.0227	0.966

α^+	-0.844(1.651)	1.974(1.662)	0.0371	1.222

Between-groups comparison of estimated parameters: BN and Healthy Controls

Results show that the group of BN patients exhibited a reduced learning rate (α) in punished trials ($p = 0.0293$, Cohen's $d = 0.958$) compared to HC (Figure 3.1). In contrast, no credible difference emerged between patients with BN and HC in the learning rate (α^+) resulted from rewarded trials ($p = 0.5207$). Individuals with BN displayed a higher decision threshold (a) ($p = 0.0002$, Cohen's d = -1.123), increased cognitive processing latency (v) for the two stimuli ($p = 0.0009$, Cohen's $d = 1.370$), and increased latency in perceiving the two stimuli and the subsequent motor-level decision (t) ($p = 0.0048$, Cohen's d = -1.314) in comparison to the healthy controls (Table 3.2). Such differences in the posterior distribution of the parameters between BN and HC are visible in the following graphs: Figure 3.3 (parameter a), 3.4 (parameter v), and 3.5 (parameter t).

Table 3.2: Posterior estimates of the parameters from the HDDMrl Model by Group (BN, HC). The learning rates (α) are presented on a logit scale. The probability (p) describes the Bayesian test comparing the posterior estimate of the parameter for the BN group versus the HC group. Standard deviations are provided in parentheses. Cohen's d indicates the effect size of the comparison between groups.

Parameter	BN Mean (SD)	HC Mean (SD)	p	Cohen's d
a	1.445 (0.296)	1.219 (0.179)	0.0002	-1.123
v	2.157 (0.682)	5.053 (0.704)	0.0009	1.370
t	1.015 (0.095)	1.183 (0.051)	0.0048	-1.314
α^-	0.485 (5.103)	0.168 (4.709)	0.0293	0.958
α^+	1.286 (1.676)	1.974 (1.662)	0.5207	-0.062

Between-groups comparison of estimated parameters: AN-R and BN

Results indicate that both groups (AN-R and BN) exhibit a general impairment in performing the PRL task; however, some differences are noted. Specifically, the AN-R group shows greater latency in stimulus perception and motor choice behavior (t) compared to BN patients ($p = 0.0406$, Cohen's $d = 0.234$), as illustrated in Figure 3.5. No credible differences in the learning rate for punished trials (α^-) are observed between the two groups ($p = 0.4926$). Similarly, no differences in the learning rate (α^+) for rewarded trials were detected ($p = 0.0795$), although a marginal trend suggests a slightly accelerated learning rate in the BN group (Figure 3.2). No credible differences were observed between the two diagnostic groups regarding cognitive processing speed (v) for the paired stimuli ($p = 0.6964$) or regarding the amount of information required to make a decision (a) between the two stimuli ($p = 0.8048$). These findings are tabulated in Table 3.3.

Table 3.3: Posterior estimates of the parameters from the HDDMrl Model by Group (AN-R, BN). The learning rates (α) are displayed on a logit scale. The probability (p) describes the Bayesian test comparing the posterior estimate of the parameter for the AN-R group against the BN group. Standard deviations are provided in parentheses. Cohen's d indicates the effect size of the comparison between groups.

Parameter	AN-R Mean (SD)	BN Mean (SD)	p	Cohen's d
a	1.469 (0.203)	1.445 (0.296)	0.8048	0.098
v	1.218 (0.663)	2.157 (0.682)	0.6964	0.303
t	0.188 (0.081)	1.015 (0.095)	0.0406	0.234
α^-	0.353 (5.407)	0.485 (5.103)	0.4926	-0.025
α^+	-0.844 (1.651)	1.286 (1.676)	0.0795	-1.283

The effect of symptom severity: A comparison between AN outpatiens and inpatients

To evaluate the impact of symptom severity on cognitive flexibility in AN, performance differences between outpatient groups (from **Study 1**) and inpatient AN-R groups during a neutral block of the PRL were examined using the HDDMrl model M6 (see Table 3.4). As observed in **Study 1**, no credible differences were found between the AN-R outpatients and HC groups in learning rates for rewarded and punished trials ($p = 0.3008$; $p = 0.0877$) in the neutral block. Similarly, parameters such as cognitive processing speed

(v), motor response time (t), and decision threshold (a) also showed no credible differences ($p = 0.1248$; $p = 0.8960$; $p = 0.5568$). However, when confronting the performances of severe AN-R and the HC group within the same block, distinct differences emerged. AN-R inpatients showed a diminished learning rate (α) in both rewarded ($p = 0.0085$, Cohen's $d = 1.487$) and punished trials ($p = 0.0102$, Cohen's $d = 1.162$). Additionally, AN-R participants presented an elevated decision threshold ($p = 0.0$, Cohen's d = -1.098), heightened cognitive processing latency (v) for both stimuli ($p = 0.050$, Cohen's d = 0.584), and an extended delay in stimulus perception leading to the subsequent motor decision (t) compared to the healthy controls ($p = 0.0$, Cohen's d = -1.812).

A comparison was conducted between the learning rates for rewarded and punished trials across the two patient groups. As illustrated in Figure 3.7, within a neutral block of the PRL task, outpatients demonstrated better performance than AN-R inpatients following a reward ($p = 0.0425$, Cohen's $d = 1.132$). No credible differences were detected in the learning rate for punished trials ($p = 0.1497$). Additionally, inpatients had a higher decision threshold ($p = 0.0$, Cohen's d = -0.881) and a prolonged delay in stimulus perception, leading to the subsequent motor decision, compared to the outpatient group ($p = 0.0$, Cohen's d = -0.650). No credible differences were observed in the drift rate parameter between the two groups ($p = 0.7077$). Overall, these findings suggest a generalized deficit in severe cases, a pattern not evident in outpatients (see Figure 3.6, 3.7, 3.8, 3.9).

Table 3.4: Posterior Estimates of the HDDMrl Model Parameters by Group (AN-R inpatients, AN-R outpatients). The learning rates (α) are displayed on a logit scale. The probability (p) describes the Bayesian test comparing the posterior estimate of the parameter for the AN-R inpatients group to that of the AN-R outpatients group. Standard deviations are provided in parentheses. Cohen's d represents the effect size measure for the group comparison.

Par	AN-inpat Mean (SD)	AN-outpat Mean (SD)	p	Cohen's d
a	1.470(0.203)	1.280(0.228)	0.0000	-0.881
v	1.237(0.793)	1.444(1.650)	0.7077	0.158
t	0.266(0.080)	0.202(0.111)	0.0000	-0.650

Par	AN-inpat Mean (SD)	AN-outpat Mean (SD)	p	Cohen's d
α^-	0.026(4.047)	2.341(3.364)	0.1497	0.625
α^+	-0.988(1.625)	1.132(2.073)	0.0425	1.132

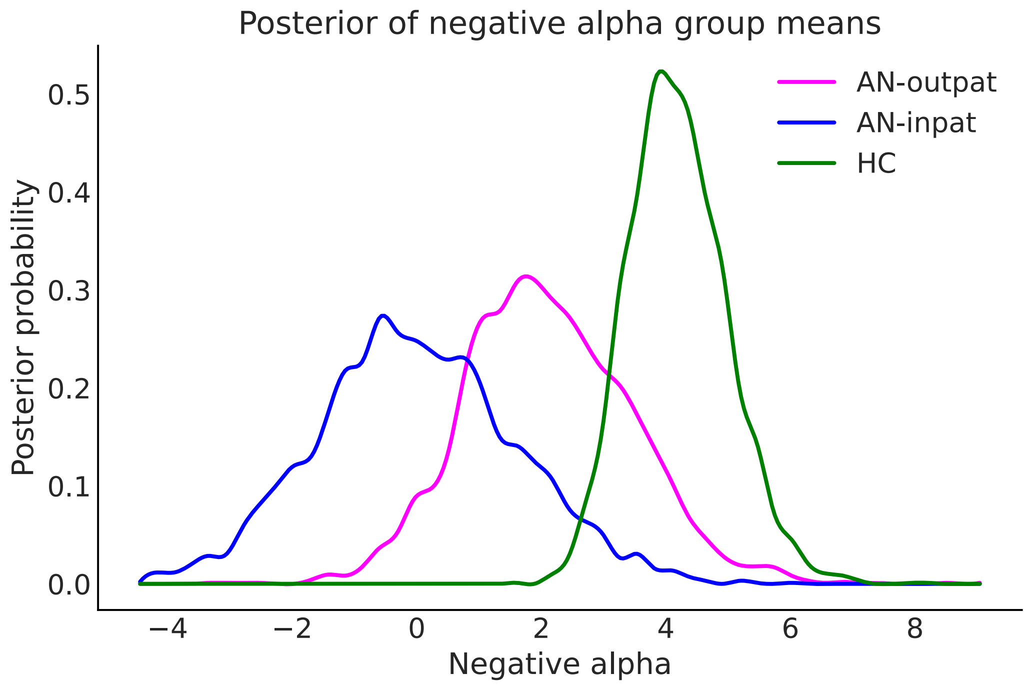

Figure 3.6: Posterior distributions comparison of the parameter α^- estimated separately for the three groups (AN-R inpatients, AN-R outpatients, HC). The pink curve represents the posterior distribution of the α^- parameter for the AN-R outpatients group; the blue curve represents the posterior distribution of the α^- parameter for the AN-R inpatients group; the green curve represents the posterior distribution of the α^- parameter for the HC group.

3.5.4 Medication effect

In light of the clinical severity observed in the patient cohort under investigation, it became imperative to examine the potential impact of medication status on cognitive flexibility. To this end, a comparative analysis employing Model M6 was conducted. This analysis compared two distinct subsets within the diagnostic group: participants on medication, who constituted 60% of the sample, and those not receiving medication. This approach aimed to discern any differential effects of medication on cognitive flexibility metrics.

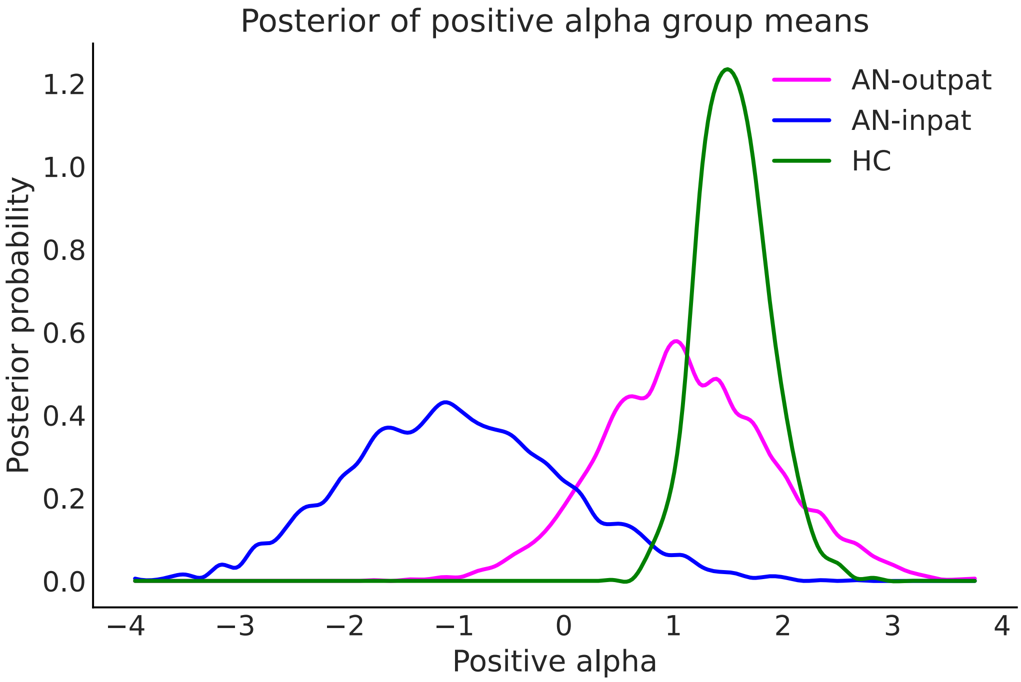

Figure 3.7: Posterior distributions comparison of the parameter α^+ estimated separately for the three groups (AN-R inpatients, AN-R outpatients, HC). The pink curve represents the posterior distribution of the α^+ parameter for the AN-R outpatients group; the blue curve represents the posterior distribution of the α^+ parameter for the AN-R inpatients group; the green curve represents the posterior distribution of the α^+ parameter for the HC group.

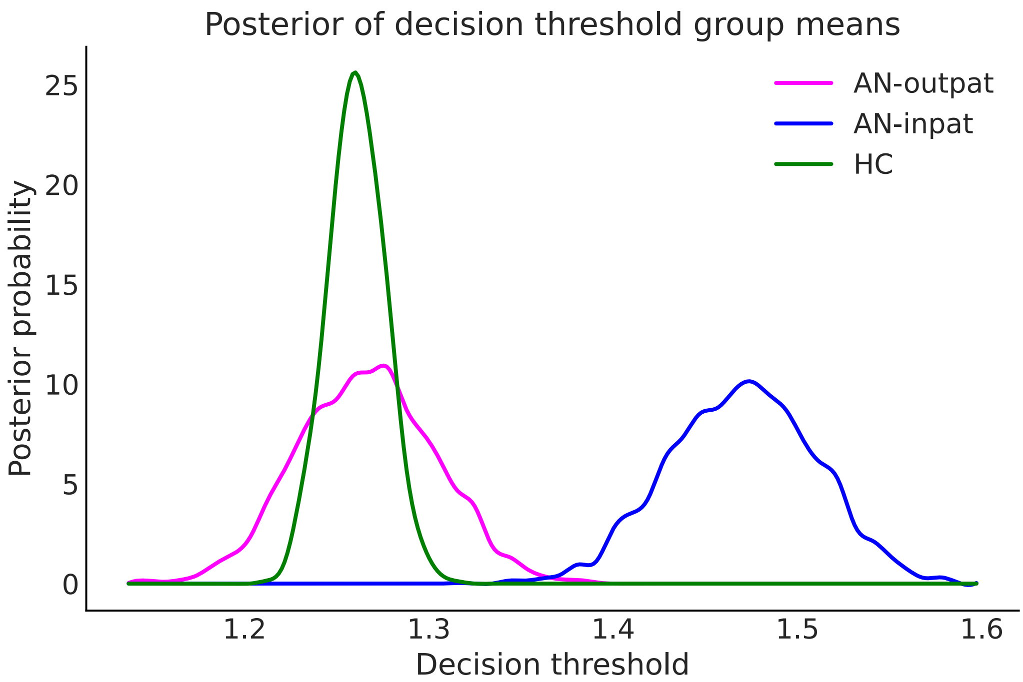

Figure 3.8: Posterior distributions comparison of the parameter a estimated separately for the three groups (AN-R inpatients, AN-R outpatients, HC). The pink curve represents the posterior distribution of the a parameter for the AN-R outpatients group; the blue curve represents the posterior distribution of the a parameter for the AN-R inpatients group; the green curve represents the posterior distribution of the a parameter for the HC group.

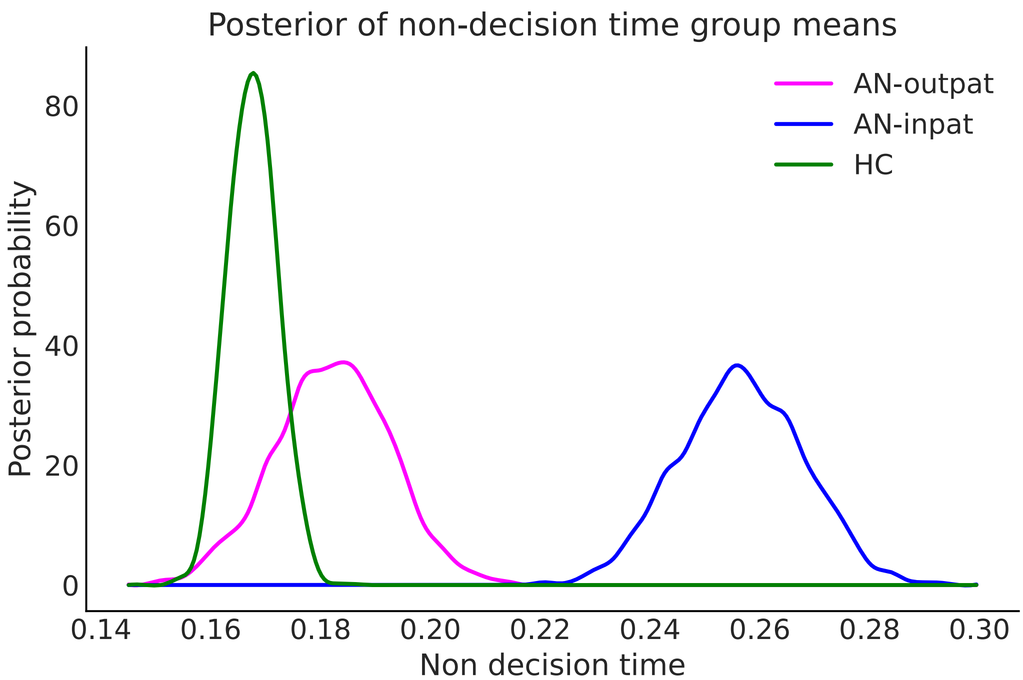

Figure 3.9: Posterior distributions comparison of the parameter t estimated separately for the three groups (AN-R inpatients, AN-R outpatients, HC). The pink curve represents the posterior distribution of the t parameter for the AN-R outpatients group; the blue curve represents the posterior distribution of the t parameter for the AN-R inpatients group; the green curve represents the posterior distribution of the t parameter for the HC group.

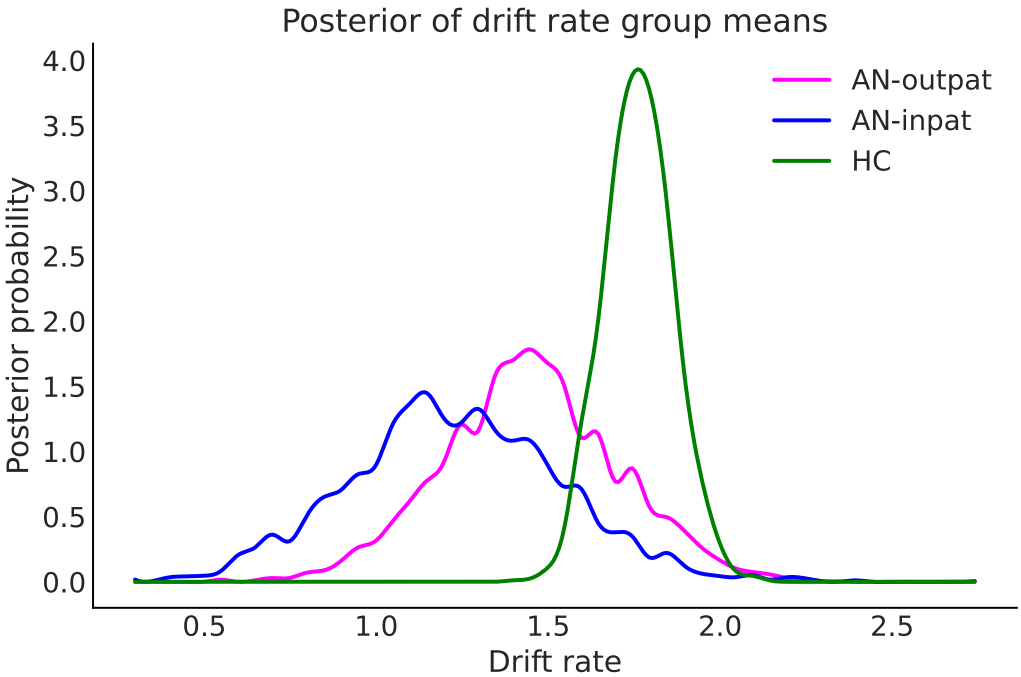

Figure 3.10: Posterior distributions comparison of the parameter v estimated separately for the three groups (AN-R inpatients, AN-R outpatients, HC). The pink curve represents the posterior distribution of the v parameter for the AN-R outpatients group; the blue curve represents the posterior distribution of the v parameter for the AN-R inpatients group; the green curve represents the posterior distribution of the v parameter for the HC group.

From the comparison of the estimated parameters, no credible differences were observed between participants who were on medication and those who were not: $\Delta\alpha^{-}$ = -1.11, 95% CI [-6.97, 4.61]; $\Delta\alpha^{+}$ = 1.56, 95% CI [-2.57, 5.96]; Δa = 0.13, 95% CI [-0.02, 0.27]; Δv = 0.19, 95% CI [-0.56, 0.96]; Δt = 0.03, 95% CI [-0.01, 0.09].

3.6 Brief Discussion

The current study sought to determine whether the cognitive flexibility deficit, previously observed in medium/low severity AN-R patients within a ED-salient context (**Study 1**), is generalized across all contexts when analyzing a group of patients with severe EDs. In other words, the research aimed to elucidate the influence of symptom severity on reinforcement learning capacities in EDs patients. The predominant hypothesis posits that as disorder severity intensifies, there is a concomitant decline in flexible adaptation capacities, pervasively affecting all aspects of an individual's life. A PRL task with neutral stimuli was administered to a group of AN-R inpatients, a group of BN inpatients, and a control group matched by age, gender, and education level. In addition to comparing the task performance of the two diagnostic groups to the HC, the two diagnostic groups were also compared. The goal of this analysis is to determine whether deficits in cognitive flexibility are present in both AN and BN, thereby supporting the concept of its *transdiagnostic* nature. This examination also aims to elucidate the unique underlying traits of these two disorders. While AN and BN are often viewed as having overlapping features (Fairburn et al., 2003), they may exhibit distinct characteristics, particularly in the realm of reinforcement learning. This distinction is crucial for understanding the specific cognitive profiles associated with each disorder. Finally, the performances of outpatients with AN-R from **Study 1** on the neutral block of the PRL were compared to the performances of the severe AN-R group on the same task, in order to directly observe the influence of symptom severity on cognitive flexibility.

Empirical Findings of the Current Investigation

From the current study, a generalized cognitive flexibility deficit emerges in inpatients with eating disorder from both diagnostic categories (BN and AN-R) compared to healthy controls, albeit with some differences. Patients with AN-R exhibited a reduced learning rate in updating their value expectations for both stimuli during reward-associated (α^+) and punishment-associated (α^-) trials when compared to healthy controls (*H1*).

Patients with BN, on the other hand, only showed a reduced learning rate in punished trials (α^-), not in rewarded trials (α^+). In rewarded trials, the speed of BN patients in updating their value expectations of the stimuli is much closer to the performance of HC (*H2*).

Patients with AN-R manifested a psycho-motor deceleration, as evidenced by parameters (v) and (t), during task performance. This was particularly evident in their cognitive processing of the two stimuli, their perceptual recognition of the stimuli, and their motor response speed, relative to the healthy controls (*H4*).

Similarly, patients with BN exhibited a deceleration in cognitive processing (parameter v), and in motor response speed (parameter t) compared to HC.

Furthermore, both AN and BN patients adopted a more cautious decision-making strategy (parameter a) in comparison to controls. Regarding the direct comparison between the two diagnostic groups, no credible differences emerge between the groups in terms of learning rate, cognitive processing speed and conservative tendencies in decision making. However, BN patients showed greater impulsivity in the speed of the motor response and greater speed in perceiving the characteristics of the two stimuli compared to the AN group (*H3*).

When comparing outpatients and inpatients with AN, a generalized deficit emerged in inpatients compared to outpatients (*H1*). Specifically, inpatients demonstrated a decreased learning rate for rewarded trials (α^+), an increased amount of information needed before choosing between the two stimuli (parameter a), and reduced perceptual and motor speed (parameter t).

Insights from the Results

In summary, findings indicate a pervasive domain-general deficit in cognitive flexibility for severe AN-R patients. Based on the HDDMrl model parameters, AN-R patients exhibited impaired performance in all facets of the decision-making process when compared to the control group. The central implication of this finding is the crucial role of symptom severity. Outpatients, under the same experimental conditions, displayed performances matched to the control group, whereas AN inpatients were notably less proficient. This pattern suggests that as the disorder progresses—with increased duration and symptom severity—the cognitive flexibility deficit becomes more pronounced. Interestingly, while the deficit in outpatients was observed only in the presence of ED-salient contexts (**Study 1**), the impairment in severe cases generalizes across contexts, regardless of its direct association with the disorder. This broader manifestation indicates a shift to a domain-general deficit in inpatients (Davis et al., 2020).

Importantly, in the present samples there was no evidence of impairment across various cognitive domains, as all participants were assessed for general cognitive abilities using the Raven's Standard Progressive Matrices. Furthermore, the influence of medication on the results was evaluated and excluded. This conclusion is supported by the comparative analysis presented in the previous section (**Medication effect**), indicating that medication use did not account for the observed results.

In contrast, patients with severe BN exhibited specific difficulties. While their performance aligned with control groups when responding to rewarded stimuli, it declines notably when confronted with symbolic punishments, bringing their performance more in line with AN patients. This differential reaction by BN patients to rewards versus punishments can be attributed to two primary mechanisms: (1) BN patients exhibited an enhanced reward sensitivity. This intensified reward perception encourages them to actively seek further rewarding experiences (Chan et al., 2014; Hagan & Forbush, 2021); (2) BN patients display a stronger emotional response to punishments and a reduced ability for cognitive reappraisal of negative emotions. Such mechanisms might negatively influence the learning processes of these patients when faced with punishing situations

(Danner et al., 2012b).

Additionally, BN patients tended to be more impulsive in their choices, especially in motor speed, consistent with the existing literature (Pearson, Wonderlich, & Smith, 2015). Such observations might further reinforce the notion that AN and BN possess a distinct underlying cognitive profile (Curzio et al., 2018). The study also provides support for the hypothesis that cognitive flexibility may serve as a *transdiagnostic* factor across psychopathologies.

The present findings emphasize the importance of distinguishing between different levels of disorder severity when investigating cognitive processes in both AN-R and BN patients (Davis et al., 2020).

Clinical Implications

The study holds significant clinical insights. Initially, it is imperative to emphasize the preliminary nature of these findings. Replication of the present findings is crucial, especially considering their potential clinical implications.

The present results suggest that cognitive flexibility deficits could serve as a maintenance mechanism for eating disorders symptoms, consistent with previous proposals (Steding et al., 2019; Walsh, 2013). The persistence of these symptoms is associated with the progressive deterioration of the patient's condition. In severe cases, the recommended therapeutic approach is hospitalization followed by residential treatment. Beyond the importance of early diagnosis, addressing the main symptoms and modifying maladaptive behaviors is crucial. Cognitive inflexibility often hinders the transformative process. From this perspective, interventions designed to reduce rigidity and encourage change could lead to faster symptom remission.

It should be emphasized that interventions aimed at enhancing cognitive flexibility, especially outside direct ED-salient contexts, appear promising for severely affected inpatients. However, as discussed in **Study 1**, for outpatients with less severe symptoms and without evident task-related difficulties (*e.g.*, the neutral block of the PRL task), the evidence remains inconclusive.

In this scenario, an example of such therapeutic intervention is the Cognitive Remediation Therapy (CRT, Davies & Tchanturia, 2005). Its efficacy in reducing cognitive and behavioral rigidity among acute AN inpatients has been documented (Tchanturia et al., 2012). CRT is designed to modify dysfunctional cognitive patterns and assist severe inpatients in overcoming cognitive inflexibility. This approach not only fosters healthier eating behaviors but also enhances therapeutic outcomes.

For severe BN patients, the results suggest potential benefits from similar treatments. However, specific elements should be adapted to accommodate the unique cognitive attributes of BN, which distinguish it from AN.

Additionally, Davis et al. (2020) proposed the potential applicability of Habit Reversal Therapy (HRT) for inpatients with an EDs. Grounded in its proven effectiveness for conditions such as Tourette's syndrome and trichotillomania (Deckersbach, Rauch, Buhlmann, & Wilhelm, 2006; Teng, Woods, & Twohig, 2006), HRT primarily targets habitual behaviors—a pronounced trait among AN-R inpatients. Its primary goal is the disruption of entrenched cue-habitual response associations. Encouragingly, prior research showcases HRT's efficacy among AN-R inpatient groups (Davis et al., 2020).

Limitations and Future Directions

1) The current study investigated cognitive flexibility by evaluating two patient cohorts: those diagnosed with AN-R and those with BN. This choice stems from the dominant focus in scientific literature on these eating disorders, with a predominant emphasis on AN. However, it is crucial to underscore that recent investigations have started to include the diagnostic category of BED, yielding compelling findings. Hence, it is posited that future research should expand the cognitive flexibility study to all Eating Disorders, while accounting for variations in severity levels within each category.

2) From the results of the present study, it is evident that differentiating symptom severity is crucial for understanding underlying mechanisms in AN-R. By discerning various symptom intensities, healthcare professionals can optimize treatment

approaches. It is reasonable to suggest that this consideration of symptom severity might be relevant across multiple psychiatric conditions. Such categorization helps researchers and clinicians better understand the nature and progression of a disorder. Future research should explore this perspective across other psychiatric disorders.

3) It should be noted that the study employs a cross-sectional design. However, a longitudinal approach may be more advantageous for investigating the evolution of a cognitive deficit throughout the progression of a disorder.

4) The comparison that was made between AN and BN was specifically and exclusively within the context of reinforcement learning processes. Reinforcement learning is a type of learning where behaviors are strengthened or weakened based on their outcomes, and this process is fundamental for cognitive flexibility, which relates to how individuals make decisions and learn from their environment. In this study, the primary objective was to investigate whether this process was impaired in both AN-R and BN, in line with the *transdiagnostic* perspective, and to identify any distinctions between the two conditions. However, in order to clarify the cognitive profiles of these two conditions, given the complexity of eating disorders, it would be interesting to extend the study to other cognitive processes. For instance, differences might emerge in terms of attention biases, decision-making strategies, or impulsivity levels. By comparing AN and BN across a broader spectrum of cognitive processes, researchers can achieve a more comprehensive understanding of the neurological and cognitive distinctions between the two disorders. Such knowledge could be highly valuable in designing more precise therapeutic interventions and advancing the overall comprehension of these conditions.

Chapter 4

Study 3: Anorexia Nervosa and Cognitive Flexibility: Examining the Impaired Component between Reversal Learning and Task-Switching

4.1 Introduction

Cognitive flexibility is a complex construct comprising distinct abilities like *set/task-switching* and *reversal learning*, which have separate neural and behavioral bases. Many studies on Anorexia Nervosa have overlooked this distinction, often using the Wisconsin Card Sorting Test (WCST) which merges these abilities and may not accurately represent cognitive flexibility. Recent research advocates for more precise tools, which specifically evaluates *reversal learning* and *set/task-switching* separately. Despite emerging evidence, many researchers continue to use generic tools, leading to inconsistent findings in AN literature. Therefore, it is crucial to use specific methodologies to accurately assess and differentiate these components for a better understanding of the underlying cognitive mechanisms that maintain the disorder and, as a consequence, for treatment implications.

4.1.1 Cognitive Flexibility: An Overview

Cognitive flexibility represents a multifaceted and complex construct, including different abilities. In particular, it has been proposed that cognitive flexibility is characterized by two main capabilities: attention *set-shifting*, the capacity to shift attention from one stimulus feature to another, also known as *task-switching* (*i.e.*, the ability to redirect attention from one cognitive task to another), and *reversal learning*, which pertains to the the faculty of readjusting stimulus-response associations based on received feedback.

From both a behavioral and neural perspectives, these abilities manifest a clear distinction. Wildes et al. (2014) highlighted that distinct neural substrates modulate *set/task-switching* and *reversal learning*. Specifically, *set/task-switching* involves the Ventromedial Prefrontal Cortex, Anterior Cingulate Cortex, and Posterior Temporal and Parietal areas. In contrast, *reversal learning* engages regions such as the Orbitofrontal Cortex and Ventral Striatum. From a behavioral point of view, Wildes et al. (2014) posited that the components of *set/task-switching* and *reversal learning* regulate distinct behavioral strategies, meaning that the two components influence and shape diverse patterns of behavior, potentially serving unique roles in adaptive responses.

Considering such evidence, it becomes imperative to employ distinct research paradigms tailored to investigate these aspects independently, rather than relying on generic methods that merge the two components. However, previous studies examining cognitive flexibility in patients with AN, did not consider this fundamental distinction, treating cognitive flexibility as a unitary construct, which often, has been identified with *set/task-switching* abilities, while rarely exclusively with *reversal learning* capabilities (Hildebrandt et al., 2015; Tchanturia et al., 2012).

4.1.2 Methodological Challenges in Assessing Cognitive Inflexibility in Anorexia Nervosa

In assessing cognitive flexibility, particularly concerning AN, various methodologies have been employed, resulting in inconsistent findings in the literature. The specific nature of the inflexibility deficit in AN remains ambiguous, potentially originating from a lack of differentiation between *set/task-switching* and *reversal learning*. Currently, it remains unclear whether AN is characterized by deficits in *set/task-switching*, *reversal learning*, or both.

4.1.2.1 Considerations around the use of Wisconsin Card Sorting Test (WCST)

The inconclusive results surrounding cognitive flexibility deficits in AN can be attributed to the widespread use of instruments like the WCST.

The WCST is the predominant cognitive flexibility task used in AN research, with nearly 50% of studies utilizing this instrument.

Despite its popularity, the WCST may not be the optimal task for examining cognitive flexibility, as it conflates the skills of *set/task-switching* and *reversal learning* (Hildebrandt et al., 2015). Moreover, its accurate execution requires engagement of additional cognitive processes, such as abstraction and conceptualization capacities, or adept attentional mechanisms (Tchanturia et al., 2012). This overlap might obscure a clear understanding of the construct under investigation.

Likewise, a recent study by Monni et al. (2023) questioned the use of WCST as the exclusive measure for assessing cognitive flexibility. The authors argued that: (1) the interpretation of WCST data is challenging, as it captures multiple dimensions simultaneously (Tchanturia et al., 2012); (2) the absence of a standardized scoring protocol for WCST exacerbates ambiguity in interpretation (Miles et al., 2021).

These considerations are supported by the inconsistent results when using the WCST for measuring cognitive flexibility in AN. For instance, a plethora of studies observed no significant differences in perseverative responses or errors on the WCST between adolescents with AN and HCs, (Andrés-Perpiña et al., 2011; McAnarney et al., 2011). Conversely, other found adolescents with AN to exhibit markedly elevated levels of perseverative errors in comparison to HCs, with pronounced effect sizes (Bischoff-Grethe et al., 2013). This ambiguity persists in research focused on adults diagnosed with AN. While several studies identified higher rates of perseverative errors among this group relative to HCs (Abbate-Daga et al., 2011; Tenconi et al., 2010), an almost equal number found no discernible difference in this metric between adults with AN and HCs (Cavedini et al., 2004; Wollenhaupt et al., 2019).

Conversely, the Reversal Learning (RL) paradigm emerges as a more precise and representative instrument for evaluating cognitive flexibility (Monni et al., 2023).

The RL paradigm specifically evaluates *reversal learning* abilities. Thus, the interpretation is less ambiguous. Through the application of computational models, it allows for the measurement of various aspects of reinforcement learning, including the capacity to update the value expectation of stimuli on a trial-by-trial basis. Monni et al. (2023) notes that the RL paradigm is more representative of daily cognitive challenges compared to WCST. It should also be observed that, while paradigms like the WCST and Task-Switching (TS) operate based on explicit and predictable rules, RL requires a robust implicit learning capability. This is due to its intrinsic nature of uncertainty and unpredictability, making it closer to the unpredictable conditions typical of real-life scenarios.

Despite these insights, many research efforts persist in using tools like WCST, often emphasizing the *set/task-switching* component while overlooking *reversal learning*.

4.1.2.2 Considerations around the use of Task-Switching (TS)

In research related to cognitive flexibility deficits in AN, the *set/task-switching* component is the most extensively examined (Wu et al., 2014). A variety of tasks are used for assessment, among which the TS task is considered ideal (Monsell, 2003). However, its application in eating disorders remains limited. The TS paradigm requires participants to shift between cognitive tasks, according to an explicit set of rule. When changing tasks, a noticeable delay and accuracy drop, known as the *switch-cost*, indicate the cognitive processes at play. Additionally, the *mix-cost* emerges when tasks are repeated in different scenarios.

In their study using the Cued Color-Shape Switching Task (CCSST), Berner et al. (2019) found that individuals with AN exhibited greater challenges in task-switching than their healthy counterparts. Specifically, those with AN demonstrated increased mix and switch costs.

On the other hand, several studies found no differences in task-switching between adolescents with AN and healthy controls (Calderoni et al., 2013, p. 2013; Hatch et al., 2010). In adult groups, while some studies, like Danner et al. (2012a), reported a decline in task-switching capabilities (notably in a study with a limited participant count), others like Spitoni, Aragonaa, Bevacqua, Cotugno, & Antonucci (2018) highlighted only delayed responses without any compromise in accuracy. However, a substantial number of studies reported no discernible differences in cognitive flexibility outcomes between adults with AN and healthy controls.

In conclusion, while some studies, like Berner et al. (2019), highlighted the challenges faced by AN individuals in task-switching, contrasting results existed in both adolescent and adult groups, indicating inconsistencies and leaving open questions regarding the true extent and nature of these deficits in AN populations.

4.1.3 Final Considerations

Which component of cognitive flexibility is compromised in AN remains unclear. However, recently the *reversal learning* dimension has received more attention in AN-related research (Bernardoni et al., 2018b; Geisler et al., 2017). So much that Hildebrandt et al. (2015) postulated that only the *reversal learning* component might be compromised in AN patients, suggesting a need for a more profound focus on this aspect in the future. The trend towards the use of non-specific paradigms, such as the WCST, coupled with the limited emphasis given to the *reversal learning* component might partially justify the inconsistency in the extant literature on the presence of a cognitive flexibility deficit in individuals with AN (Miles et al., 2020). The question around the optimal way to examine cognitive flexibility in eating disorders, particularly in AN, is not merely theoretical but holds significant clinical implications.

Cognitive flexibility has been pointed out as a potential treatment target in AN. Nonetheless, the efficacy of interventions may be compromised if the employed assessment tools for cognitive flexibility are imprecise or inadequate. Hagan et al. (2020) underscore the imperative to consolidate a unified approach for the examination of cognitive flexibility, especially in the context of AN.

The increasing urgency is highlighted by the diverse array of assessment tools cited in the literature, including instruments such as the WCST, the TS, the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997), the Trail Making Test (Reitan, 1958), the CatBat (Tchanturia et al., 2005), the Intra-and Extra-Dimensional Task (ID/ED, Robbins et al., 1998), tasks from the Delis-Kaplan Executive Function System (D-KEFS, Delis, Kaplan, & Kramer, 2001), and others (Miles et al., 2020). Although each of these tests aims to evaluate cognitive flexibility, the extensive range of tools has yielded results that are frequently contradictory. The dilemma is not just in the variety of instruments but also in the lack of clarity regarding the specific cognitive flexibility aspect being examined. Numerous investigations, in their attempt to assess cognitive flexibility, have neglected to specify the exact component under examination (Caudek et al., 2021). This complicates the comparison of results between various studies. In conclusion, for a comprehensive understanding of the role of cognitive flexibility in AN, it is imperative to adopt methodologies that distinctly evaluates the components of cognitive flexibility (*set/task-switching* and *reversal learning*), using instruments expressly designed for such assessments (Wildes et al., 2014).

In the current study, the aim is to discern which component of cognitive flexibility is impaired in AN and contribute to the persistence of the disorder. Based on previous research, it is hypothesized that the *reversal learning* component may be the primary area of concern.

4.2 The Present Study

The aim of the present study is to examine which dimension of cognitive flexibility is compromised in patients with AN-R, with a specific focus on *reversal learning* abilities in comparison to *set/task-switching* capabilities. To address this question, a series of computerized tasks were administered to a group of patients with Anorexia Nervosa Restrictive type (AN-R) and a healthy control group: The Probabilistic Reversal Learning (PRL) Task to measure *reversal learning* abilities, the Task-Switching (TS) to evaluate *set/task-switching* capabilities, and finally, the Wisconsin Card Sorting Test (WCST), given its widespread use in literature (Hildebrandt et al., 2015). Subsequent analyses will compare the performances in each of these tasks between the AN-R patient group and the healthy control group.

Existing research suggests that AN-R patients may have impaired *reversal learning* abilities relative to healthy controls. In contrast, their *set/task-switching* abilities, as assessed by TS and WCST, might remain unaffected.

4.3 Method

4.3.1 Participants

The overall sample consisted of 37 outpatient diagnosed with AN-R ($M_{age} = 20.17$, $SD_{age} = 5.36$) and 192 HC ($M_{age} = 20.34$, $SD_{age} = 2.11$), matched for age, gender, and education level. Both groups included female participants. The patients were recruited from an Italian center specializing in the treatment of eating disorders, namely the Specchidacqua Institute in Montecatini (Pistoia, Italy). All recruited patients met the criteria for an eating disorder diagnosis as per DSM-5 (American Psychiatric Association, 2013). Diagnostic assessment was conducted by professionals, psychiatrists, and psychologists affiliated with the aforementioned institution. Data collection was conducted approximately 2 weeks after the patients' initial intake and completion of initial evaluations. Besides the primary diagnosis of an eating disorder, the presence of psychiatric comorbidities was also considered, identified during the diagnostic process through additional psychological assessments by mental health professionals affiliated with the center. However, no comorbidities were detected in the patient group. Notably, 5% of the patients were on medication. The same professionals evaluated the eligibility criteria for the participation of patients in the research using semi-structured clinical interviews. Exclusion criteria included the presence of neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis.

The HC group was recruited on a voluntary basis through advertisement on social media or at the university. Subsequently, the HCs were administered the Eating Attitudes Test-26 (EAT-26) screening test to assess any tendencies towards developing an eating disorder. Participants who scored above the cut-off (≥ 20) on the EAT-26 were deemed at risk for developing eating disorders (Dotti & Lazzari, 1998) and were excluded from the research. Exclusion criteria for the HC group also included anomalous values of BMI, currently undergoing treatment or having a diagnosis of an eating disorder, neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis.

All participants reported normative cognitive functioning, as assessed by the Standard Progressive Matrices test by Raven (Raven, 2003). The majority of participants were Caucasian. All participants were right-handed, demonstrated competency in writing/reading and understanding the Italian language, and were not informed about the objectives of the study.

4.3.2 Materials

All experiments were programmed and executed using the online platform Psytoolkit (https://www.psytoolkit.org).

Probabilistic Reversal Learning Task (PRL)

In the present investigation, a PRL task incorporating domain-specific stimuli (*e.g.*, a slice of cake) was employed. The rationale for utilizing a block with ED-salient stimuli originates from two primary considerations: (1) In **Study 1**, our findings indicated that individuals diagnosed with medium/low severity AN-R manifested cognitive flexibility impairments specifically when exposed to ED-salient stimuli. Consequently, reiterating a comparison between a neutral condition and a context-dependent condition within the PRL framework would be superfluous. (2) The primary goal of this research focuses on clarifying the distinctions between two comparable cognitive processes that have been historically confused and imprecisely assessed in prior literature.

Stimuli were selected from the IAPS database (Lang & Bradley, 2007): five images of objects (IAPS #7010, 7020, 7034, 7056, 7170) and five images of food (IAPS #7461, 7260, 7470, 7451, 7405). The PRL block involved the presentation of two images per trial (randomly selected from the aforementioned categories), for a total of 160 trials. These 160 trials are split into 4 epochs of 40 trials each. In every epoch, one of the two stimuli is rewarded 70% of the time, while the other 30%. Upon transitioning to a subsequent epoch, the stimulus-response contingency is reversed; the previously majorly rewarded stimulus now becomes rewarded 30% of the time, and the previously majorly penalized stimulus becomes rewarded 70% of the time (see Figure 2.1). In this case as well, the reward is symbolized by an image of a symbolic euro coin, whereas the punishment is represented by an image of a crossed-out symbolic euro coin.

Task-Switching (TS)

The TS, as described by Monsell (2003), requires participants to continuously shift between different tasks that increase in difficulty (Figure 4.1). Each transition from one task to another is clearly announced and accompanied by instructions specific to the upcoming task. The TS task comprises three sub-tasks with 40 trials each. In this version, the screen displays four panels. During each trial, two images appear in one of these panels, each image representing a specific category. For example, one image might belong to the "food" category while the other to the "plant" category. Within these categories there are two modalities: the "food" category contains high and low-caloric food images, and the "plant" category features images of flowers and leaves.

Participants were instructed to judge only one of the two stimulus categories presented, which varied across the three sub-tasks. In the first sub-task, stimuli appeared in the upper panels, and participants judged the food-related images using different keys, depending on whether the images were of high or low-caloric foods. The plant category was ignored. Conversely, in the second sub-task, stimuli appeared in the lower panels, and participants judged the plant images, distinguishing between flowers and leaves, while overlooking the food category. In these tasks, there was no category switch or shift. Only the first task involved a switch, resulting from a transition to a different task. When the task changed, the stimulus category to be judged changed accordingly. However, the third sub-task introduced frequent switches. Images could appear in either the upper or lower panels, requiring participants to adjust their judgment criteria based on the panel's location. If images were in the upper panel, the "food" category was judged; if in the lower panel, the "plant" category became the focus.

Mistakes were followed by a penalty signal. The key point of interest in this task is the transition or *switching* between tasks. A delay in reaction times is expected immediately after a *switch*.

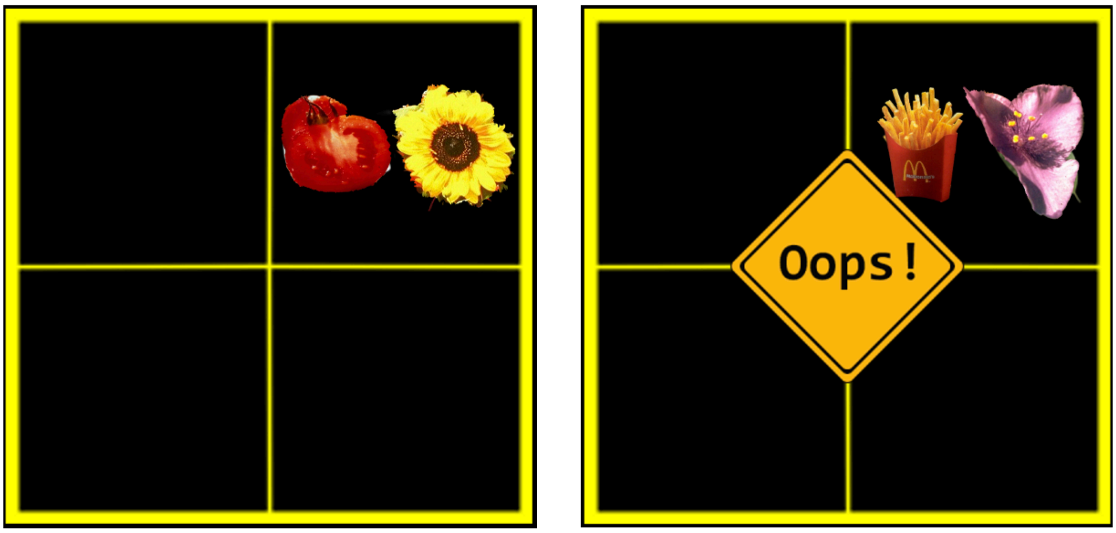

Figure 4.1: This figure illustrates the Task-Switching version used in this study. In the upper panels, two stimuli are presented: one showcases a low-caloric food paired with a neutral item, and the other displays a high-caloric food alongside a neutral item. A correct response from the participant triggers a pleasing sound and nothing appears on the screen (left image), while an incorrect response results in an error signal and an accompanying unpleasant sound (right image). The stimuli were taken from the IAPS database.

Wisconsin Card Sorting Test

Participants also completed a computerized version of the WCST, given its widespread use in literature to measure cognitive flexibility (Grant & Berg, 1993). In the WCST, four cards from a specific category are presented (e.g., a red triangle, two green stars, three yellow crosses, and four blue circles) alongside a target card. The participant must match the target image to one of the four category cards. The cards display a varying number of colored symbols, which can be circles, triangles, crosses, or stars, and these can appear in one of four distinct colors, including red, green, blue, and yellow. The number of symbols depicted on each card can range from one to four. After every choice that matches the target card to one of the four fixed cards, positive or negative feedback is provided to indicate whether the given response is right or wrong (*Correct* or *Incorrect*). Unpredictably during the task, the rule by which cards are matched changes.

In the version administered in this study, the task was characterized by six blocks of 10 trials each. Participants were not informed beforehand neither about the rule to follow in order to make the correct stimulus choice, nor that the rule would change after a certain number of trials. Reaction time was measured in milliseconds, with a maximum response time of 10 seconds.

There are two types of possible errors: perseverative errors, where the participant persists in a wrong rule, and non-perseverative errors. The objective is to make as few perseverative errors as possible. In the traditional data analysis, it is essential to focus on the analysis of reaction times and perseverative errors.

4.3.3 Procedure

The data collection involved three phases. In the first phase, inclusion and exclusion criteria were examined to select participants for the healthy control group. Those who voluntarily chose to participate in the study as control subjects were asked to provide their weight and height to calculate the BMI, and to complete the EAT-26 screening test. Participants surpassing the cut-off threshold (≥ 20) were excluded from the final sample. Out of the initial 213 participants in the control group, 21 were excluded due to scores above the cut-off on the EAT-26 test. Subsequently, a further evaluation was conducted to exclude the presence of neurological disorders, suicidal ideation, drug or alcohol addiction, and psychosis. Patients underwent a similar assessment by mental health professionals at the facility from which they were recruited. In the second phase, all participants completed the Raven's Standard Progressive Matrices (Raven, 2003) to evaluate their general cognitive functioning. Lastly, in the third phase, participants executed the Probabilistic Reversal Learning task, the Task-Switching task, and the Wisconsin Card Sorting Test. Each phase of the study took approximately 30 minutes on average. Participation was voluntary, and no incentives were offered to participants. The study was conducted in adherence to the Declaration of Helsinki and was approved by the Ethics Research Committee of the University of Florence (Prot. No. 0178082). All participants provided their informed consent, received a privacy disclosure, and agreed to partake in the research.

4.4 Data Analysis

Data analysis was conducted using Python software (https://www.python.org) and R version 4.2.1 (https://cran.r-project.org/index.html).

As in the previous studies (**Study 1** and **Study 2**), PRL data were analyzed by estimating the 6 parameters of the computational Hierarchical Reinforcement Learning Drift Diffusion Model (HDDMrl, Pedersen & Frank, 2020). The *Task-Switching* data were analyzed using only the portion of the HDDMrl model that estimates the parameters of the Drift Diffusion Model (HDDM). The WCST task was analyzed through the computational Weighted Parallel Reinforcement-Learning (wP-RL) model by Steinke, Lange, & Kopp (2020).

Comparison of Estimated Parameters. Parameters were estimated separately for each group and task, and then compared across groups. Three distinct group comparisons were performed: (1) Between the AN-R patient group and the control group using the PRL task with food stimuli, (2) between the AN-R patient group and the control group using the Task-Switching task with food stimuli, and (3) between the AN-R patient group and the control group using the classic version of the WCST task

Classification. Lastly, the discriminatory capacities of the PRL and TS instruments between the clinical and control groups were assessed. These tasks are directly associated with the *reversal learning* and *set/task-switching* facets, respectively. To evaluate their classification capacity, ROC curve plots (*i.e.*, visual representations of classification model performance) were generated for both instruments, followed by a report of the AUC index for each tool.

wP-RL Model for WCST Data Analysis

Traditional analysis of WCST data has primarily focused on perseverative errors as a key behavioral index. However, the introduction of Reinforcement Learning computational models offers a fresh perspective on WCST analysis. Specifically, the wP-RL model differentiates between model-based and model-free decision-making strategies, a pivotal distinction for comprehending the nuances of decision-making. Model-based reinforcement learning predicts positive feedback after using a rule, guiding decisions towards specific criteria like color, shape, or number. Based on actual feedback, these predictions adjust to influence future choices. Conversely, the model-free reinforcement learning operates independently of the rule and choice criteria, relying instead on previously learned stimulus-response associations (Steinke et al., 2020), which form the basis for habitual behaviors.

In the wP-RL model, feedback expectations are integrated linearly, shaping the probability of certain responses. Subsequently, the model identifies the following parameters:

- Model-based learning rate after positive feedback (α^+_{MB}): The learning rate for the model-based reinforcement learning strategy in trials where a reward is given.
- Model-based learning rate after negative feedback (α^-_{MB}): The learning rate for the model-based reinforcement learning strategy in trials where a penalty is given.
- Model-free learning rate after positive feedback (α^+_{MF}): The learning rate for the

model-free reinforcement learning strategy in trials where a reward is given.

- Model-free learning rate after negative feedback (α^{-}_{MF}): The learning rate for the model-free reinforcement learning strategy in trials where a penalty is given.
- Model-based inertia γ_{MB}: Quantifies the impact of expectations to receive feedback in previous trials on the current response.
- Model-free inertia γ_{MF}: Quantifies the impact of expectations to receive feedback in previous trials on the current response.
- Temperature (t): The measure to which the choices made align with current feedback expectations, thus aiming to maximize gain.
- Weighting of model-based and model-free RL (w): A trade-off between the two reinforcement learning strategies of model-based and model-free. Specifically, it quantifies the relative strength of model-based reinforcement learning compared to model-free.

4.5 Results

4.5.1 Preliminary Analysis

In the current study, consistent with the previous investigations (**Study 1** and **Study 2**), a quality check was conducted on the PRL, TS, and WCST data. Subjects who passed the quality check totaled N = 221, of which N = 32 were patients with AN-R, and N = 189 were healthy subjects.

Additionally, the same analysis as in the previous studies was conducted for the selection of the HDDMrl and DDM models (for details see Appendix C).

4.5.2 Results from the HDDMrl and DDM Models

Both the HDDMrl model (M6) and the DDM model (M7) were estimated using 15,000 iterations, with a burn-in of 5,000 iterations. The $\hat{R}$ values for all the estimated parameters were below 1.1, indicating good model convergence.

To investigate which of the two components of cognitive flexibility is compromised in patients with AN-R, posterior estimates of the HDDMrl model parameters for the PRL task were compared between the two groups. Simultaneously, posterior estimates of the DDM model parameters for the Task Switching task were contrasted between the two groups involved in the study. In conclusion, a classification analysis was conducted with the goal of evaluating the efficacy of both tools in differentiating the clinical group from the control group.

Comparison between AN-R and Healthy Controls in the PRL Task

The estimated parameters of the HDDMrl model were compared between the two groups. Specifically, a Bayesian general linear model was applied (estimated using the MCMC *sampling* method with 4 chains of 4000 interactions and a *warmup* of 2000) to predict parameter values based on affiliation in either the clinical group or the control group. Concerning the parameter α, the results indicate that the AN-R patient group exhibits a reduced learning rate in trials where a reward is given (α^+ parameter: b = 1.26, 95% CI [0.40, 1.91], Bayesian effect size = -0.24, pr[$b > 0$] = 0.99) and in punished trials (α^- parameter: b = 0.88, 95% CI [0.50, 1.43], Bayesian effect size = -0.90, pr[$b > 0$] = 0.99) compared to healthy controls (Figure 4.2; Figure 4.3). Furthermore, individuals with AN-R demonstrated slower cognitive processing (v) of the two stimuli (v parameter: b = 0.45, 95% CI [0.27, 0.65], Bayesian effect size = -0.55, pr[$b > 0$] = 0.99) and required a greater amount of information to make their choice (a parameter: b = -0.07, 95% CI [-0.13, -0.01], Bayesian effect size = 0.08, pr[$b < 0$] = 0.98) - Figure 4.4 and Figure 4.5. Conversely, no credible differences emerged between the AN-R patient group and the healthy control group in the perceptual speed of the two stimuli and the subsequent motor-level choice (t), as the credibility interval includes zero (Table 4.1).

Table 4.1: Posterior estimates from the Bayesian General Linear Model for predicting the values of the HDDMrl model parameters based on affiliation to either the AN patient group or the HC group. Cohen's d denotes the effect size of the between-group comparison.

Par.	Group effect PE(95%CI)	Cohen's d
a	-0.07 (-0.13, -0.01)	0.08

Par.	Group effect PE(95%CI)	Cohen's d
v	0.45 (0.27, 0.65)	-0.55
t	0.01 (-0.01, 0.03)	0.08
α^-	0.88 (0.50, 1.43)	-0.90
α^+	1.26 (0.40, 1.91)	-0.24

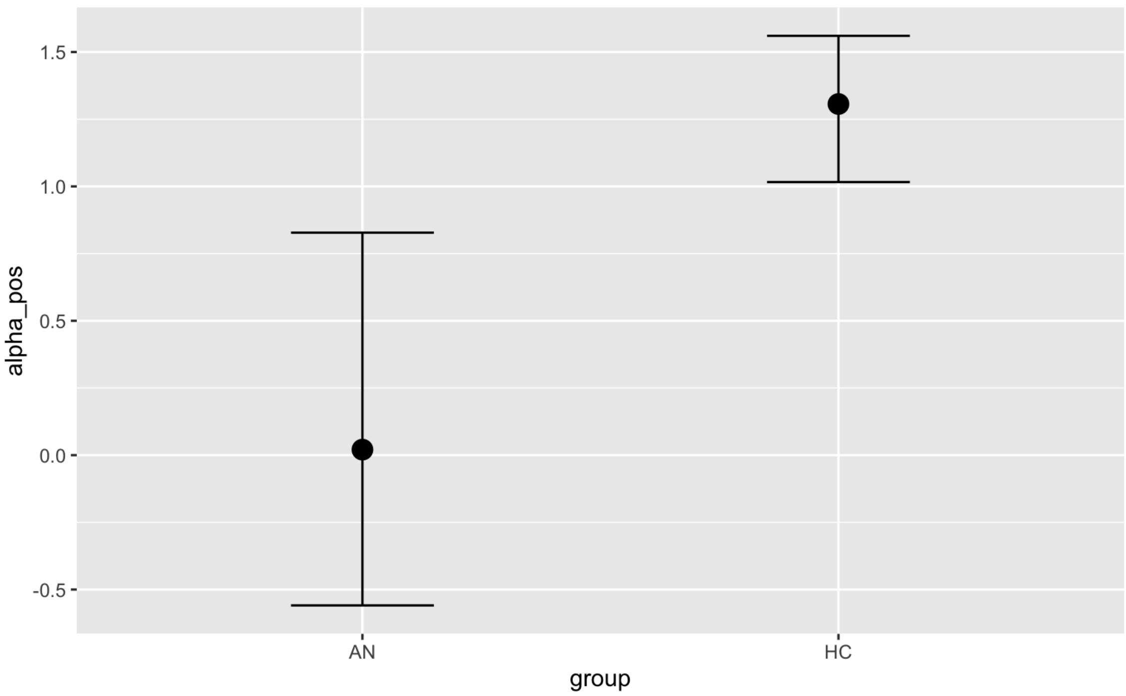

Figure 4.2: Mean of the posterior distribution for α^+ parameter from the HDDMrl model as a function of *group* in the PRL task. Vertical bars denote credible intervals.

Comparison between AN-R and Healthy Controls in the TS Task

The results reveal that there are no credible differences between the group of patients with AN-R and the HC group in performing the Task-Switching task. Specifically, there are no differences in the speed at which participants cognitively process the characteristics of the stimulus in the various trials (parameter v), nor in motor and perceptual speed (parameter t), nor in the amount of information needed to make their choice (parameter a). Indeed, for all estimated parameters of the model, the credibility interval of the comparison between the two groups includes zero (Table 4.2).

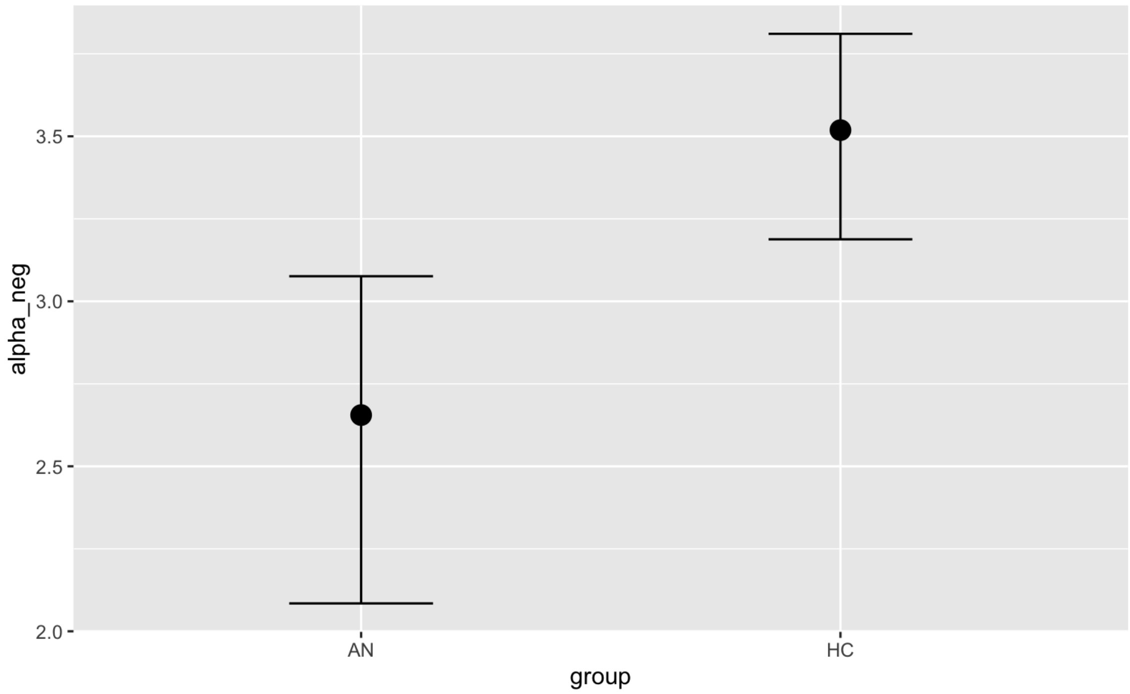

Figure 4.3: Mean of the posterior distribution for α^- parameter from the HDDMrl model as a function of *group* in the PRL task. Vertical bars denote credible intervals.

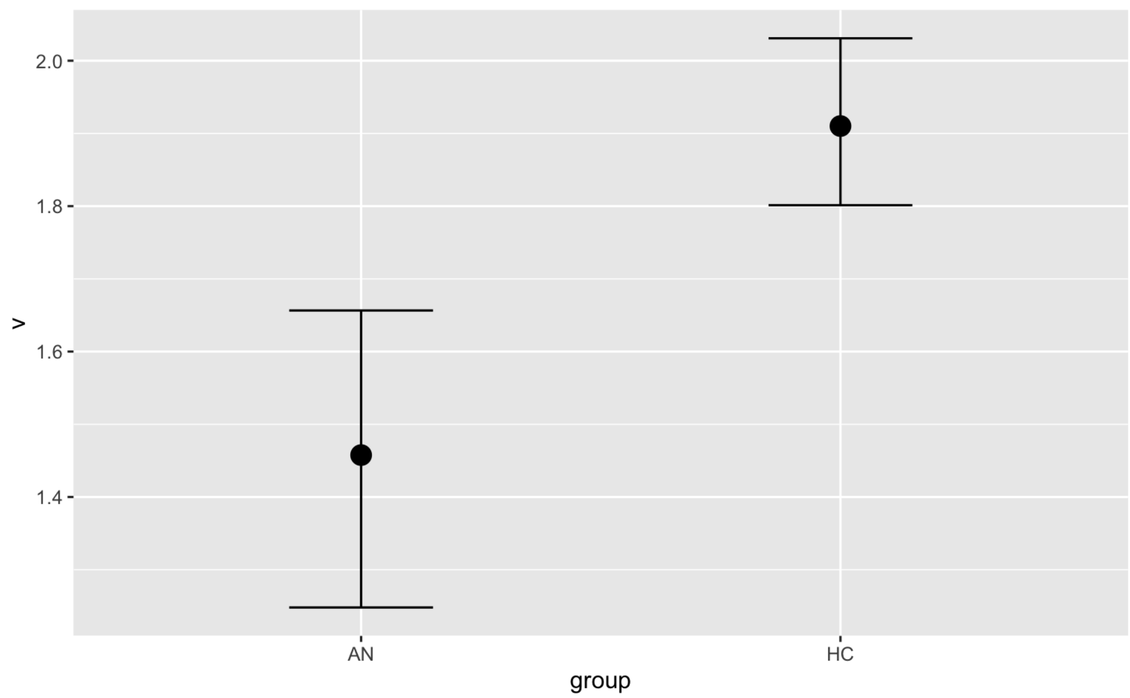

Figure 4.4: Mean of the posterior distribution for v parameter from the HDDMrl model as a function of *group* in the PRL task. Vertical bars denote credible intervals.

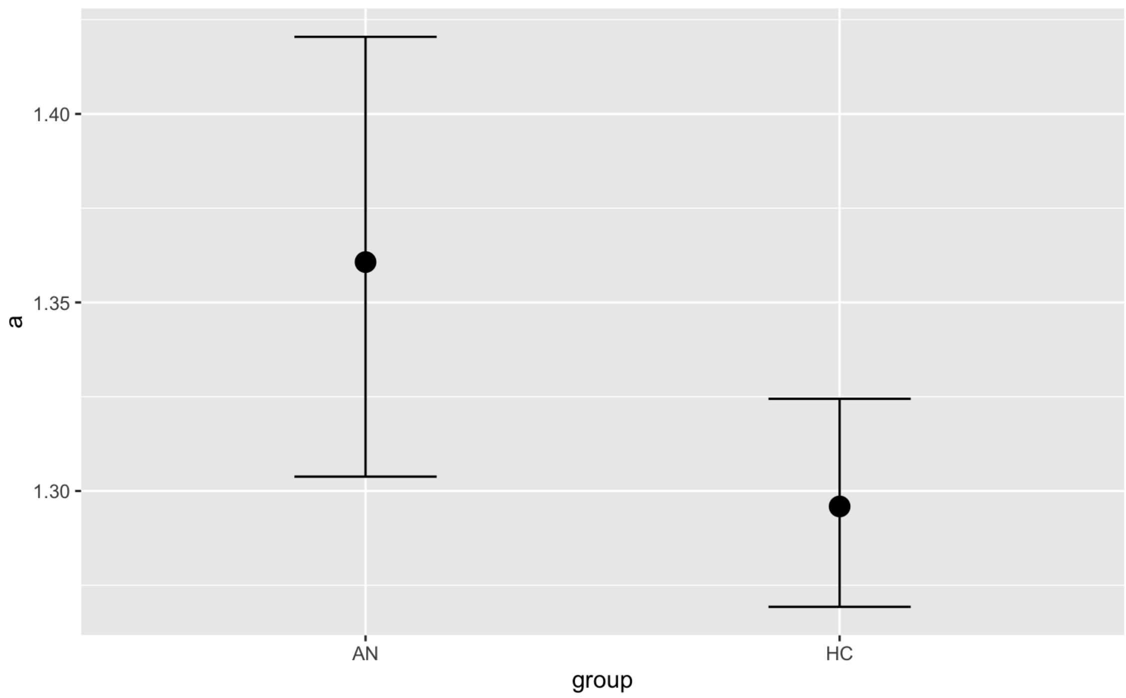

Figure 4.5: Mean of the posterior distribution for a parameter from the HDDMrl model as a function of *group* in the PRL task. Vertical bars denote credible intervals.

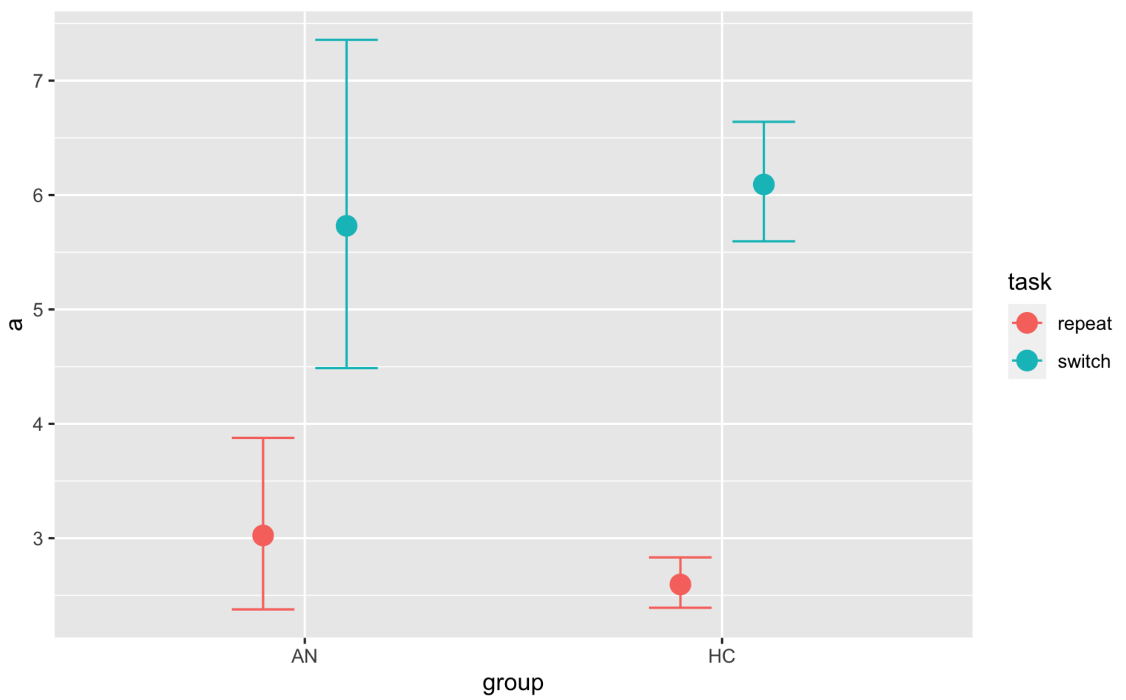

Figure 4.6: Mean of the posterior distribution for a parameter from the HDDMrl model as a function of *group* and *task* in the TS task. Vertical bars denote credible intervals.

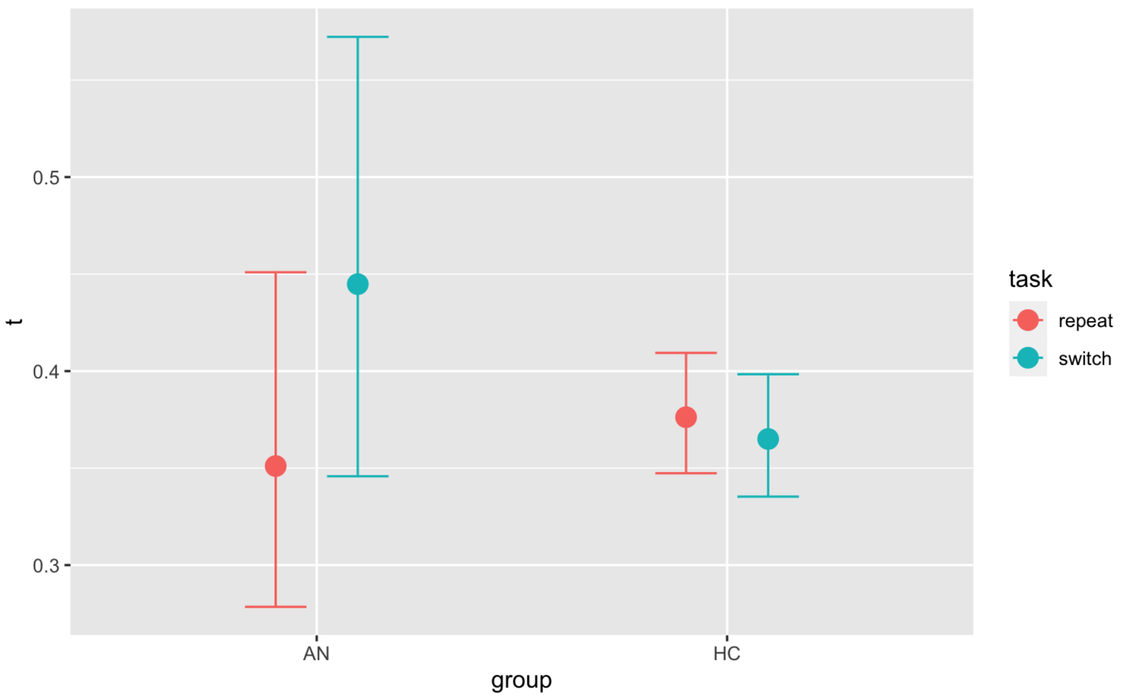

Figure 4.7: Mean of the posterior distribution for t parameter from the HDDMrl model as a function of *group* and *task* in the TS task. Vertical bars denote credible intervals.

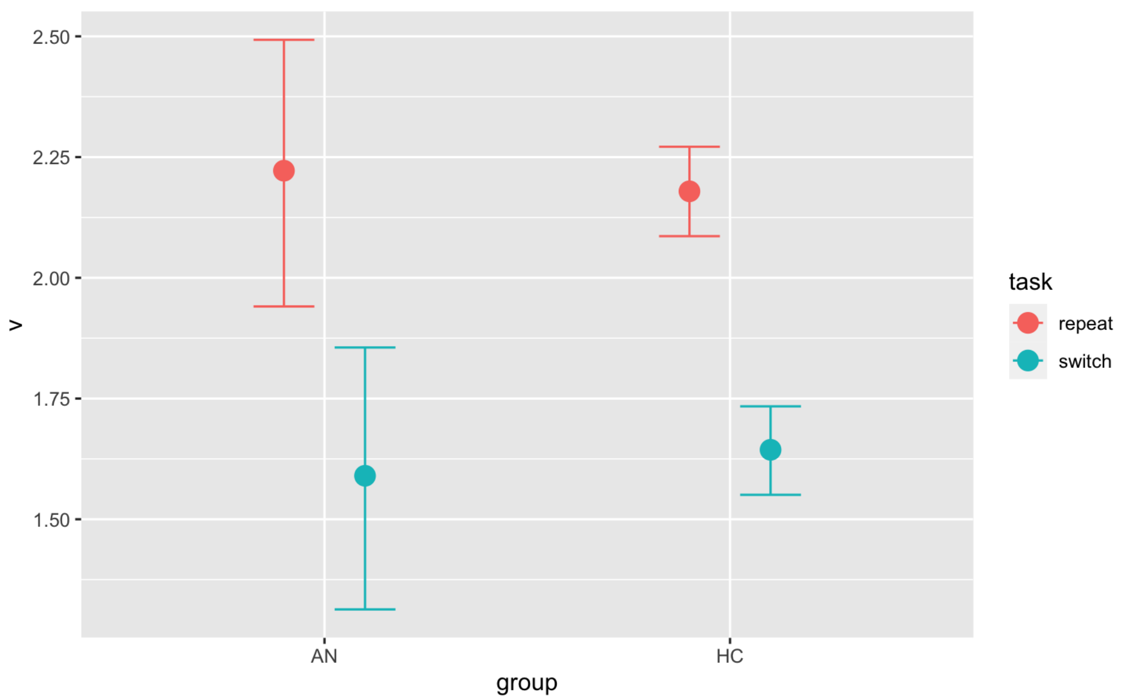

Figure 4.8: Mean of the posterior distribution for v parameter from the HDDMrl model as a function of *group* and *task* in the TS task. Vertical bars denote credible intervals.

However, the *switch* effect, a reflection of the cognitive process involved in changing from one task to another, was evident within both groups individually. Specifically, for the HC group and the AN-R patients, the effect is pronounced for parameter a and v (a parameter: b = 0.64, 95% CI [0.28, 0.98], Bayesian effect size = 0.39; v parameter: b = -0.63, 95% CI [-0.93, -0.33], Bayesian effect size = 0.19). This indicates that both groups were actively engaging in the task-switching mechanism (Berner et al., 2019; Monsell, 2003). Yet, when examining the magnitude or characteristics of this switch effect between AN and HC, no differences arise. In essence, while both groups show a clear *switch* effect when looked at separately, the manner and extent to which they experience this effect is notably similar - Figures 4.6, 4.7, 4.8.

Table 4.2: Posterior estimates from the Bayesian General Linear Model for predicting the values of the DDM model parameters based on affiliation to either the AN patient group or the HC group. Cohen's d denotes the effect size of the between-group comparison.

Par.	Group effect PE(95%CI)	TS effect PE(95%CI)	Inter. PE(95% CI)	Cohen's d
a	-0.15 (-0.42, 0.10)	0.64 (0.28, 0.98)	0.22 (-0.17, 0.60)	0.39
v	-0.04 (-0.32, 0.24)	-0.63 (-0.93, -0.33)	0.10 (-0.22, 0.41)	0.19
t	0.07 (-0.18, 0.31)	0.23 (-0.11, 0.58)	-0.26 (-0.63, 0.10)	-0.24

Considerations on the WCST

For the WCST, a series of parameters were estimated using the wP-RL model, as detailed in Steinke et al. (2020). The model estimates 8 parameters (α^+_{MB}, α^-_{MB}, α^+_{MF}, α^-_{MF}, γ_{MB}, γ_{MF}, t, w) capturing the *decision-making* processes underlying task performance. No credible differences emerged between the group of patients with AN-R and the healthy control group for any of these parameters.

This suggests that performance on the WCST is similar for both groups, with neither the clinical group nor the control group showing compromised performance (see Appendix D).

Classification Analysis for the PRL and TS Tasks

The Receiver Operating Characteristic (ROC) curve is a pivotal graphical representation for evaluating and comparing the discriminatory abilities of classification models or instruments. It plots sensitivity (true positive rate) against 1-specificity (false positive rate). The area under the curve (AUC) provides an overarching measure of a model's ability to distinguish between true positive and false positive classes.

In this study, the classification abilities of two instruments, PRL and TS, are assessed. The PRL exhibits an AUC of 0.80, suggesting a strong discriminatory ability (Figure 4.9). Broadly speaking, this implies that the PRL possesses an 80% probability of correctly classifying a true positive case over a false positive case. Thus, the PRL is a proficient instrument for differentiating between AN-R patients and healthy controls.

On the other hand, the TS exhibits an AUC of 0.53, which is only slightly above the threshold of random classification (an AUC of 0.5 indicates random classification capability, similar to flipping a coin). Its ability to differentiate between the clinical and control groups appears limited (Figure 4.10).

As a result, while the PRL stands out as a robust classification tool in this sample, the TS exhibits limited discriminating capacity and might not be considered reliable for classification tasks in clinical or research practices.

4.6 Brief Discussion

The present study aims to determine which facet of cognitive flexibility, *set/task-switching* or *reversal learning*, is compromised in AN-R. Accurate and specialized measuring tools were employed for each of the aforementioned components. Three specific tasks were administered to both a group of AN-R diagnosed patients and a control group of healthy participants: the PRL task, the Task-Switching, and the WCST. The performances of both groups on the PRL and Task-Switching tasks were subsequently compared to discern potential impairments in *set/task-switching*, *reversal learning*, or both, among the AN-R patients. Additionally, given its widespread use in

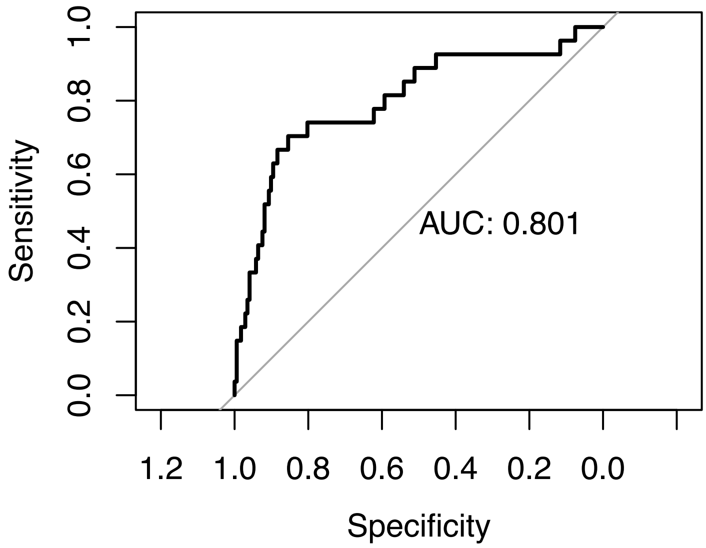

Figure 4.9: Plot of the ROC Curve for PRL in predicting Anorexia Nervosa diagnosis. The AUC serves as an indicator of the model capacity to differentiate between classes. Here, the AUC is 0.80, suggesting a strong ability of the model to accurately classify subjects.

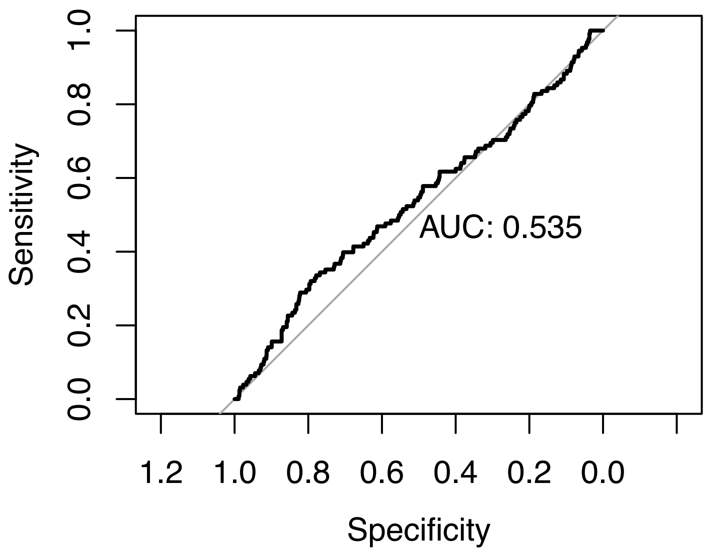

Figure 4.10: Plot of the ROC Curve for Task-Switching in predicting Anorexia Nervosa diagnosis. The AUC serves as an indicator of the model capacity to differentiate between classes. Here, with an AUC of 0.535, the discriminative power is only marginally better than random chance.

the literature, the WCST performance was particularly examined to detect its ability to measure cognitive flexibility in patients with AN-R.

Empirical Findings of the Current Investigation

The results indicate a pronounced deficit in the *reversal learning* component in AN-R patients compared to the healthy controls. In contrast, there were no discernible differences in performance on the Task-Switching and WCST tasks between AN-R patients and the healthy control group.

Specifically, in the PRL task, AN-R patients showed a significantly reduced speed in updating their value expectations for the two stimuli, both in rewarded trials (α^{+}) and in punished trials (α^{-}) compared to the control group. Additionally, these patients required more information to make a decision (a) and were slower in cognitive processing of the two stimuli (v). However, there was no notable difference in the motor or perceptual speed with which the task was performed (t).

Regarding the *Task-Switching* task, there were no performance differences between patients and controls in psycho-motor speed of task execution (v and t) nor in the amount of information needed to make a choice (a). This suggests that the *set/task-switching* component is intact in AN-R patients.

Furthermore, there were no performance differences in the WCST between the two groups. This implies that, in this specific context, the WCST might not be a reliable measure of cognitive flexibility.

In the study, the PRL task, specifically adapted to assess the *reversal learning* component, demonstrated efficacy in differentiating between patients and control groups. In contrast, the TS task, intended to evaluate the *set/task-switching* component, was found to be unreliable for this differentiation.

The Wisconsin Card Sorting Test

In the current study, the classic version of the WCST was employed to measure cognitive flexibility. The WCST is frequently referenced in literature as a primary tool for assessing cognitive flexibility. However, recent research underscores its limitations. It has been highlighted that the WCST is not solely sensitive to cognitive flexibility; performance can also be influenced by other cognitive domains, such as working memory and attention (Tchanturia et al., 2012). Moreover, the presence of reinforcement learning mechanisms in the WCST could confound the clear assessment of *set/task-switching* abilities. As identified by Hildebrandt et al. (2015), the WCST captures both *reversal learning* and *set/task-switching*, leading to potentially overlapping data. Given its intricate nature, various interpretations and scoring methods have been proposed and implemented in AN research, adding an element of complexity (Miles et al., 2021). A recent shift towards computational modeling, as suggested by Haynos et al. (2022), offers a more precise method for analyzing data derived from learning-related tasks. Based on this consideration, Steinke et al. (2020) emphasized that computational modeling could address the challenge of scoring variability inherent to the WCST. For this reason, in the present study, a sophisticated computational model was employed for WCST data analysis. Such models are considered a highly effective way to achieve the greatest possible precision and accuracy in results (Haynos et al., 2022).

Despite ongoing debates about the efficacy of the WCST in assessing cognitive flexibility in AN, it continues to be a prominent instrument in this field of research.

The objective of the current study revolves around identifying which facet of cognitive flexibility is most compromised in AN and determining the optimal instrument for this assessment. Given the widespread recognition of the WCST, its capacity to measure cognitive flexibility was compared with more specialized tasks, the TS and the PRL.

In the specific cohort analyzed in this study, the WCST did not provide a precise measure of cognitive flexibility, despite the use of sophisticated methods for data analysis. Notably, no differences were observed between AN-R and HC performances on the WCST. Conversely, such variance was evident in the PRL task, suggesting a discernible disparity in cognitive flexibility between AN and HC that the WCST failed to detect.

The Task-Switching

In the present study, a domain-specific version of the TS paradigm was employed to enhance the robustness of our findings, particularly building upon insights from **Study 1**.

The TS has been designed specifically to evaluate the *set/task-switching* dimension of cognitive flexibility (Monsell, 2003). Over time, numerous studies have explored *set/task-switching* abilities in AN using a range of tools. Often, this skill is conflated with cognitive flexibility, neglecting the *reversal learning* component. The TS paradigm is particularly suitable as it does not incorporate reinforcement learning mechanisms, unlike other tasks, such as the WCST, used for assessing *set/task-switching* in AN.

With the governing rule in TS being explicit, actions are not driven by feedback but by the inherent ability to shift attention between cognitive tasks. This typically necessitates a transition from one feature of a stimulus to another. Conversely, tasks like the WCST find it challenging to separate efficient switching from learning and are potentially influenced by sensitivities to rewards and punishments, elements known to be compromised in AN (Foerde & Steinglass, 2017). In the TS approach, actions depend on the actual engagement of the *set/task-switching* ability.

The objective was to identify this specific component in AN patients. Nevertheless, the body of research on *set/task-switching* in AN has yielded mixed results. While some studies have reported impairments in *set/task-switching* capabilities, others have found no such deficits (Miles et al., 2020). These disparities might stem from the use of varied tools, some of which might introduce confounding factors when measuring *set/task-switching* in AN. However, when using a more specific tool like the TS, as in this study, which excludes reinforcement learning mechanisms and other cognitive processes, the results become clearer: there is no evident impairment in *set/task-switching* in AN compared to HC. This suggests that the *set/task-switching* capability may remain intact in AN.

It is worth noting, however, that the examined sample consisted of AN outpatients of the restrictive type with low to medium symptom severity. Given this context, findings suggest that the reinforcement learning component of cognitive flexibility, rather than the *set/task-switching*, may be the impaired mechanism, in this specific population. Further investigations are recommended.

Insights from the Results

In summary, the findings are consistent with recent literature, emphasizing the importance of examining more closely the *reversal learning* component of cognitive flexibility in Eating Disorders (Hildebrandt et al., 2015). This component could play a significant, though not necessarily exclusive, role in the dysfunctional patterns observed in such individuals. Abilities associated with *reversal learning* are crucial for individuals to adapt to environmental changes and to embrace goal-oriented behaviors. This mechanism allows individuals to re-evaluate and adjust behaviors when they are no longer congruent or adaptive. AN-R patients appear to have a marked deficit in this adaptive capacity (Foerde et al., 2021; Walsh, 2013). Conversely, *set/task-switching* ability seems preserved in AN-R patients. This might explain the inconsistent findings in literature, where the heterogeneity of tools used and the ambiguity regarding the specific dimensions of cognitive flexibility explored have led to contrasting results.

The study underscores that, in AN patients, the compromised component of cognitive flexibility is the *reversal learning* ability.

Moreover, the research emphasizes potential limitations in using the WCST as an accurate tool to evaluate cognitive flexibility in AN patients, despite its widespread use in scientific literature (Miles et al., 2020).

Clinical Implications

The results from the current study hold potential significant implications for both assessment and therapeutic intervention in AN. However, it is crucial to underscore the preliminary nature of these results. Rigorous replications are needed, given the implications these findings could present for shaping future, targeted therapeutic interventions. Cognitive inflexibility may serve as a maintenance factor in AN, thereby hindering the success of therapeutic interventions. To optimize therapeutic strategies, a more profound understanding of this cognitive deficit is crucial. This is supported by two salient points:

1) The extensive variety of assessment tools employed in prior research to measure cognitive inflexibility has created significant challenges in accurately evaluating cognitive rigidity in AN patients and in assessing the effectiveness of interventions. For instance, Hagan et al. (2020) highlights the difficulty in measuring the efficacy of interventions such as Cognitive Remediation Therapy (CRT) in targeting inflexibility in AN, attributed primarily to the methodological variability in assessments across studies. Such heterogeneity in methodologies complicates the synthesis of coherent conclusions regarding the effectiveness of CRT in improving cognitive flexibility deficits in AN.

2) *Set/task-switching* and *reversal learning* represents two distinct cognitive functions. Consequently, therapeutic interventions or training programs tailored to enhance each of these abilities would necessitate unique approaches, given the fundamental differences in their underlying neural and cognitive mechanisms.

The current study represents an initial attempt to address the existing gaps in knowledge. Its primary goal was to clarify the specific component of cognitive flexibility that is impaired in AN patients, and consequently, to identify the most suitable assessment tool for detecting such deficits. Nevertheless, it remains crucial that these findings necessitate rigorous replication for validation in future research. If future research confirms the dual nature of cognitive flexibility, interventions will need to be specific in targeting those distinct abilities, rather than employing a generic approach.

Limitations and Future Directions

1) As in the case of **Study 1**, the AN-R patient group can be considered of medium/low severity. Future studies should expand the investigation of various components of cognitive flexibility among patients with a spectrum of severity levels.

2) The distinct nature of cognitive flexibility components was tested in this specific sample, characterized by patients diagnosed with AN-R. Future research should

extend this investigation to other diagnostic categories, including BN and BED.

3) The current study underscored the limitations of WCST in evaluating cognitive flexibility among AN-R patients. However, the WCST comes with various scoring methods (Monni et al., 2023) and has seen several recent adaptations (Steinke et al., 2020). Subsequent investigations should further explore the efficacy of WCST in assessing cognitive flexibility.

4) It is crucial to acknowledge that this study is based on a cross-sectional design. Yet, for a deeper and more effective investigation into the nature of cognitive flexibility in eating disorders, adopting a longitudinal approach may be more beneficial. Such an approach enables a comprehensive analysis of how various aspects of cognitive flexibility are affected by factors like age differences, the severity of symptoms, and their changes over time.

Chapter 5

General Discussion and Conclusions

Cognitive inflexibility (CI) is proposed as a *transdiagnostic* underlying mechanism that drive the onset and maintenance of various psychopathologies. Such a *transdiagnostic* factor refers to a diverse array of mechanisms (*e.g.,* cognitive, behavioral, emotional, etc.) present in a wide range of diagnostic categories. These mechanisms shape observed symptoms and are crucial for the onset and persistence of disorders (Visu-Petra & Mărcuș, 2019). CI, indicative of an inability to adapt to changing environmental demands, potentially intensifies symptoms and behavioral dysfunctions. Hence, it is a significant research focus across many mental health conditions.

Recently, there has been a growing emphasis on targeting *transdiagnostic* factors to enhance treatment outcomes. Such factors can perpetuate disorders and, in turn, reduce intervention efficacy (Kozak & Cuthbert, 2016). A persistent challenge in clinical psychology is adopting a more effective perspective on psychopathology. This involves accurately classifying mental disorders and identify their root causes. Current classification systems like the DSM-5-TR (American Psychiatric Association, 2022) and International Classification of Disease-11 (ICD-11, World Health Organization, 2018) offer limited guidance in predicting and addressing mental health issues. Despite extensive research, the etiology of mental disorders, including their genetic and neural bases, is still mostly unknown.

Challenges in diagnostic classification stem from limited understanding of the root causes of mental disorders. Comprehensive insights into these origins are pivotal for refining mental health condition classifications. However, progress in this field is often hindered by limitations inherent in traditional classification systems.

In light of these considerations, a paradigm shift has been proposed: departing from traditional systems like DSM and ICD, which primarily focus on symptoms, and instead targeting on basic *transdiagnostic* factors for deeper mental disorder understanding. The RDoC initiative, presented by Kozak & Cuthbert (2016), provides a holistic perspective on mental disorders, merging insights from biology, psychology, neuroscience, and genetics. Its core objective is discerning and addressing the fundamental mechanisms that cause and maintain these disorders. Importantly, these mechanisms are often *transdiagnostic*, suggesting their presence across a range of psychopathologies. By exploring not just the symptoms but also the root processes that sustain mental disorders, a clearer understanding of their etiology can be achieved, leading to more personalized and effective treatment approaches.

The RDoC introduces a refreshing approach, emphasizing the importance of understanding the etiology of mental disorders. By organizing mental functions into domains and constructs, RDoC supports a *transdiagnostic* perspective, shifting focus from traditional diagnostic categories to universal underlying processes.

For instance, the Cognitive System domain of RDoC includes the Cognitive Control construct. This covers goal-oriented behaviors, adaptability to environmental demands, inhibition of inappropriate actions, and engagement in habitual behaviors when required—essentially, the core cognitive areas this research addresses. Furthermore, the Negative and Positive Valence System domains cover the Reinforcement Learning mechanisms investigated here using the Probabilistic Reversal Learning task (Garcia-Burgos, 2022).

From the RDoC lens, cognitive flexibility could act as a *transdiagnostic* factor across various disorders, highlighting the necessity to study it comprehensively. Adopting this perspective could transform our understanding of psychopathology, leading to pivotal shifts in treatment strategies—a true paradigm shift (Michelini, Palumbo, DeYoung, Latzman, & Kotov, 2021).

In the current study, the aim was to explore CI as a *transdiagnostic* factor within the framework of the RDoC perspective, with a primary focus on Eating Disorders. The disorders of primary interest were Anorexia Nervosa (AN), and to a lesser extent, Bulimia Nervosa (BN). Traditional symptom-based classifications for Eating Disorders present several challenges, such as varied symptom presentations under the same diagnosis, overlapping criteria, and fluidity between diagnoses. Furthermore, a significant overlap of symptoms is often observed across different diagnostic categories. These complexities can obstruct effective treatment strategies and complicate outcome predictions. Given these challenges, a shift in perspective appears necessary.

Although CI has been frequently associated with Eating Disorders, especially AN, existing literature provides inconsistent results. To date, the role of CI as a potential underlying factor for the development and maintenance of AN remains an open question. The inconsistencies in previous studies might arise from both theoretical and methodological challenges:

(1) In examining CI, many studies have focused on neurological dysfunctions, raising questions about its true nature. Is CI a domain-general trait impaired across various domains, or is it specific to certain contexts? Hitchcock et al. (2022) highlights the importance of an integrative approach that merges neuroscientific insights with behavioral and contextual observations.

(2) Many studies overlook variations in patient characteristics, notably different levels of symptom severity, which can substantially influence outcomes related to the study of CI (Tchanturia et al., 2012).

(3) Cognitive flexibility is a multifaceted construct, characterized by *reversal learning* and *set/task-switching* abilities. This complexity led to a notable variance in the instruments employed across studies to measure CI. This heterogeneity in tools can result in disparate findings and hinder cross-study comparisons. A call for a more unified measurement methodology is evident in Hagan et al. (2020).

(4) The variability in data analysis techniques and scoring methodologies across studies impedes straightforward comparisons. Whereas numerous investigations have centered on basic behavioral markers, there is a growing recognition of the need to transition to sophisticated computational models for a deeper insight into the phenomenon (Haynos et al., 2022).

The aim of the present research was to clarify the role of CI in AN, particularly focusing on the Restrictive type (AN-R), which is most closely associated with rigidity. Additionally, BN was also examined for CI, albeit to a lesser extent. This research represents an initial step towards broadening the study of CI to other psychological disorders. By addressing limitations identified in previous studies, the aspiration is to extend the investigation of CI to various psychopathologies.

In order to achieve this goal, three separate investigations were undertaken. Each study aimed to address a specific challenge posed by previous research. In this manner, a more detailed understanding of AN-R emerges. It became clear that AN-R is marked by a pronounced deficit in cognitive flexibility, which may contribute to sustaining the disorder. However, its potential role as a risk factor for the development of eating disorders becomes less clear when examining the performance of an at-risk group. In this study design, no learning deficits were evident in the at-risk group, suggesting that CI may not serve as a risk factor for the onset of eating disorders. Therefore, while it appears to be a maintaining factor, its role in the onset remains inconclusive, highlighting the need for further research. Interestingly, the observed cognitive flexibility deficit appeared to be domain-specific in patients with low to medium severity of AN-R. Such individuals exhibited adaptability in response to general environmental cues (**Study 1**). However, this adaptability diminished in scenarios linked to ED-salient stimuli, leading them to revert to deeply rooted, dysfunctional behaviors. According to Walsh's model, this could be attributed to their reliance on stimulus-response associations that had previously resulted in positive feedback, trapping them into entrenched habits, when confronted with stressful situations (Walsh, 2013). As a result, their goal-oriented actions— those driven by specific intended outcomes or goal— often get neglected by these habits. Yet, in neutral contexts, the efficacy of both behavioral types appears equal (Foerde et al., 2021).

On delving deeper into patients with severe AN-R, the cognitive flexibility deficit was found to increase, assuming a domain-general characteristic (**Study 2**). This shift implies that these patients largely default to habitual behaviors, overcoming goal-oriented actions. It remains a topic of discussion whether the exacerbation of symptoms leads to increased cognitive rigidity, or if persistent cognitive rigidity contributes to the worsening of symptoms (Davis et al., 2020). Furthermore, in this study, the *transdiagnostic* nature of CI is supported by a comparative analysis of two distinct diagnostic groups, AN-R and BN. In both diagnostic categories, a deficit in cognitive flexibility was observed, albeit with some variations. These differences highlight the unique aspects of AN-R and BN as separate disorders, supporting the *transdiagnostic* idea of CI.

Interestingly, these findings predominantly arose from examining the *reversal learning* component of cognitive flexibility. When the focus shifted to *set/task-switching*, using methods like the Task-Switching task and the widely-recognized Wisconsin Card Sorting Test (WCST), no discernible impairments emerged (**Study 3**).

According with this evidence, it is plausible to suggest that the variability in past findings may have resulted from an incorrect perspective on the issue, which led to the use of imprecise evaluation techniques. The current research distinguishes itself by providing a innovative and comprehensive examination of cognitive inflexibility in AN. This examination originated from a conventional debate regarding the nature of cognitive deficits: are they domain-general or domain-specific? By applying this debate to the study of CI in AN, it was possible to elucidate the nature of CI (Hitchcock et al., 2022). It appears to be domain-specific in outpatients with milder symptoms. Given the multifaceted nature of AN, which presents with varying degrees of severity and types of symptoms, it was essential to consider these variations in disease severity to understand their impact on cognitive processes. Through this lens, cognitive inflexibility was found to be domain-general in cases of greater disease severity. Additionally, by examining CI in BN, the study supports the concept of *transdiagnostic* cognitive processes. Finally, by delving deeper into the two facets of cognitive inflexibility (*reversal learning* and *set/task-switching*), especially emphasizing the role of the WCST, this study highlights the need for precise and customized assessment strategies in future research.

It is important to note that while providing valuable insights, all the presented studies have a significant limitation due to their cross-sectional design. Future research would benefit from a longitudinal approach, enabling a more comprehensive understanding of cognitive flexibility as a *transdiagnostic* maintaining factor in eating disorders. Observing changes over time, as opposed to relying solely on cross-sectional data, would not only deepen this understanding but also potentially clarify the role of CI as a causative risk factor in the onset of eating disorders—a topic that remains unresolved in the current study and continues to be an active area of research.

That being said, findings highlight the importance of accurately assessing CI to achieve clear results and promote consistency across studies.

However, it is essential to underscore the need for replicating these findings in subsequent studies, perhaps moving from a cross-sectional design to a longitudinal approach. Should these results be validated, treatment strategies would need to be refined to effectively address *transdiagnostic* factors.

The findings from this study have potential clinical implications. Within AN, a deficit in cognitive flexibility might impede treatments focused on promoting functional behaviors and modifying maladaptive patterns, like Cognitive-Behavioral Therapy (CBT, Fairburn et al., 2003). Even though CBT and family-based approaches are seen as the most effective interventions for AN, their success rates stand below 50% (Kass, Kolko, & Wilfley, 2013). It is plausible that without addressing CI, encouraging adaptive behaviors could pose significant challenges. However, to discern the need for interventions targeting this inflexibility, a precise evaluation of its presence and nature is crucial. In the case of BN, a parallel scenario is observed, with CBT treatments demonstrating approximately a 50% success rate (Kass et al., 2013). In this context as well, the presence of CI could potentially impede the efficacy of the treatment. Building on this premise, this research puts forth the hypothesis that the evaluation of CI in AN, should employ specialized instruments, such as the PRL, to gauge the *reversal learning* aspect of CI. The degree of symptom severity should also be taken into account, paired with an assessment for either domain-general or domain-specific deficits in cognitive flexibility.

The implications drawn from this research might also be applicable to other psychopathologies. Addressing *transdiagnostic* components is pivotal for improving treatment outcomes, as these elements can amplify disorders and hinder the success of interventions across various mental conditions. Tailoring treatments to target specific mechanisms identified within the RDoC framework might promote the development of more effective and individualized therapeutic strategies. A primary insight from this research emphasizes the value of the RDoC framework in providing a comprehensive understanding of psychopathological complexities.

In summary, the research suggests that CI might serve as a *transdiagnostic* maintaining factor in both AN and potentially in BN. Individuals affected by these disorders exhibit rigid cognitive and behavioral patterns, which tend to intensify as the disorder progresses. This rigidity is predominantly associated with the *reversal learning* aspect of cognitive flexibility. These insights could be extended to explore other mental disorders associated with CI, particularly in cases where existing findings are either inconclusive or insufficient to challenge the current understanding.

References

Abbate-Daga, G., Buzzichelli, S., Amianto, F., Rocca, G., Marzola, E., McClintock, S. M., & Fassino, S. (2011). Cognitive flexibility in verbal and nonverbal domains and decision making in anorexia nervosa patients: A pilot study. *BMC Psychiatry*, *11*, 1–8.

Adoue, C., Jaussent, I., Olié, E., Beziat, S., Van den Eynde, F., Courtet, P., & Guillaume, S. (2015). A further assessment of decision-making in anorexia nervosa. *European Psychiatry*, *30*, 121–127.

Aloi, M., Rania, M., Caroleo, M., Bruni, A., Palmieri, A., Cauteruccio, M. A., ... Segura-Garcia, C. (2015). Decision making, central coherence and set-shifting: A comparison between binge eating disorder, anorexia nervosa and healthy controls. *BMC Psychiatry*, *15*, 1–10.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., Text Revision). Arlington, VA: American Psychiatric Publishing.

Andrés-Perpiña, S., Lozano-Serra, E., Puig, O., Lera-Miguel, S., Lázaro, L., & Castro-Fornieles, J. (2011). Clinical and biological correlates of adolescent anorexia nervosa with impaired cognitive profile. *European Child & Adolescent Psychiatry*, *20*, 541–549.

Atwood, M. E., & Friedman, A. (2020). A systematic review of enhanced cognitive behavioral therapy (CBT-e) for eating disorders. *International Journal of Eating Disorders*, *53*, 311–330.

Banca, P., Harrison, N. A., Voon, V., & Brand, M. (2016). *Compulsivity Across the*

Pathological Misuse of Drug and Non-Drug Rewards. *10*, 1–12.

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.

Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment*. University of Pennsylvania Press.

Ben-Zion, Z., Fine, N. B., Keynan, N. J., Admon, R., Green, N., Halevi, M., et al.others. (2018). Cognitive flexibility predicts PTSD symptoms: Observational and interventional studies. *Frontiers in Psychiatry*, *9*, 477.

Bernardoni, F., Geisler, D., King, J. A., Javadi, A.-H., Ritschel, F., Murr, J., ... Ehrlich, S. (2018b). Altered medial frontal feedback learning signals in Anorexia Nervosa. *Biological Psychiatry*, *83*, 235–243.

Bernardoni, F., King, J. A., Geisler, D., Birkenstock, J., Tam, F. I., Weidner, K., ... Ehrlich, S. (2018a). Nutritional status affects cortical folding: Lessons learned from Anorexia Nervosa. *Biological Psychiatry*, *84*, 692–701.

Bernardoni, F., King, J. A., Geisler, D., Ritschel, F., Schwoebel, S., Reiter, A. M., ... 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*, 736–747.

Berner, L. A., Romero, E. M., Reilly, E. E., Lavender, J. M., Kaye, W. H., & Wierenga, C. E. (2019). Task-switching inefficiencies in currently ill, but not remitted anorexia nervosa. *International Journal of Eating Disorders*, *52*, 1316–1321.

Bischoff-Grethe, A., McCurdy, D., Grenesko-Stevens, E., Irvine, L. E. Z., Wagner, A., Yau, W.-Y. W., et al.others. (2013). Altered brain response to reward and punishment in adolescents with anorexia nervosa. *Psychiatry Research: Neuroimaging*, *214*, 331–340.

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.

Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the trail making test. *Nature Protocols*, *1*, 2277–2281.

Brooks, S., Prince, A., Stahl, D., Campbell, I. C., & Treasure, J. (2011). A systematic review and meta-analysis of cognitive bias to food stimuli in people with disordered eating behaviour. *Clinical Psychology Review*, *31*, 37–51.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*.

Calderoni, S., Muratori, F., Leggero, C., Narzisi, A., Apicella, F., Balottin, U., ... Urgesi, C. (2013). Neuropsychological functioning in children and adolescents with restrictive-type anorexia nervosa: An in-depth investigation with NEPSY–II. *Journal of Clinical and Experimental Neuropsychology*, *35*, 167–179.

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*, 317–328.

Caudek, C., Sica, C., Marchetti, I., Colpizzi, I., & Stendardi, D. (2020). Cognitive inflexibility specificity for individuals with high levels of obsessive-compulsive symptoms. *Journal of Behavioral and Cognitive Therapy*.

Cavedini, P., Bassi, T., Ubbiali, A., Casolari, A., Giordani, S., Zorzi, C., & Bellodi, L. (2004). Neuropsychological investigation of decision-making in anorexia nervosa. *Psychiatry Research*, *127*, 259–266.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, *37*, 715.

Chamberlain, S. R., Fineberg, N. A., Blackwell, A. D., Robbins, T. W., & Sahakian, B. J. (2006). Motor inhibition and cognitive flexibility in obsessive-compulsive disorder and trichotillomania. *American Journal of Psychiatry*, *163*, 1282–1284.

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*, 157–167.

Chang, P. G., Delgadillo, J., & Waller, G. (2021). Early response to psychological treatment for eating disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *86*, 102032.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*, 4563–4567.

Cury, M. E. G., Berberian, A., Scarpato, B. S., Kerr-Gaffney, J., Santos, F. H., & Claudino, A. M. (2020). Scrutinizing domains of executive function in binge eating disorder: A systematic review and meta-analysis. *Frontiers in Psychiatry*, *11*, 288.

Curzio, O., Maestro, S., Rossi, G., Calderoni, S., Giombini, L., Scardigli, S., ... Muratori, F. (2018). Transdiagnostic vs. Disorder-focused perspective in children and adolescents with eating disorders: Findings from a large multisite exploratory study. *European Psychiatry*, *49*, 81–93.

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*, 571–578.

Danner, U. N., Evers, C., Stok, F. M., Elburg, A. A. van, & Ridder, D. T. de. (2012b). A double burden: Emotional eating and lack of cognitive reappraisal in eating disordered women. *European Eating Disorders Review*, *20*, 490–495.

Danner, U. N., Sanders, N., Smeets, P. A., Meer, F. van, Adan, R. A., Hoek, H. W., & Elburg, A. A. van. (2012a). Neuropsychological weaknesses in anorexia nervosa: Set-shifting, central coherence, and decision making in currently ill and recovered women. *International Journal of Eating Disorders*, *45*, 685–694.

Darcy, A. M., Fitzpatrick, K. K., Colborn, D., Manasse, S., Datta, N., Aspen, V., ... Lock, J. (2012). Set-shifting among adolescents with bulimic spectrum eating disorders. *Psychosomatic Medicine*, *74*, 869–872.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078.

Davies, H., & Tchanturia, K. (2005). Cognitive remediation therapy as an intervention for acute anorexia nervosa: A case report. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association*, *13*, 311–316.

Davis, L., Walsh, B. T., Schebendach, J., Glasofer, D. R., & Steinglass, J. E. (2020). Habits are stronger with longer duration of illness and greater severity in anorexia nervosa. *International Journal of Eating Disorders*, *53*, 683–689.

Deckersbach, T., Rauch, S., Buhlmann, U., & Wilhelm, S. (2006). Habit reversal versus supportive psychotherapy in tourette's disorder: A randomized controlled trial and predictors of treatment response. *Behaviour Research and Therapy*, *44*, 1079–1090.

Delis, D., Kaplan, E., & Kramer, J. (2001). *Examiner's Manual: Delis-Kaplan Executive Functioning Systems (D-KEFS)*. San Antonio, TX: The Psychological Corporation.

den Ouden, H. E. M., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., ... Cools, R. (2013). Dissociable effects of dopamine and serotonin on reversal learning. *Neuron*, *80*, 1090–1100.

Diamond, A. (2005). Attention-deficit disorder (attention-deficit/hyperactivity disorder without hyperactivity): A neurobiologically and behaviorally distinct disorder from attention-deficit/hyperactivity disorder (with hyperactivity). *Development and Psychopathology*, *17*, 807–825.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Diamond, A. (2016). *Why improving and assessing executive functions early in life is critical.*

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

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

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders*, *16*, 363–370.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy*, *41*, 509–528.

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*, 463–481.

Foerde, K., & Steinglass, J. E. (2017). Decreased feedback learning in anorexia nervosa persists after weight restoration. *International Journal of Eating Disorders*, *50*, 415–423.

Forney, K. J., Buchman-Schmitt, J. M., Keel, P. K., & Frank, G. K. (2016). The medical complications associated with purging. *International Journal of Eating Disorders*, *49*, 249–259.

Fradkin, I., Strauss, A. Y., Pereg, M., & Huppert, J. D. (2018). Rigidly applied rules? Revisiting inflexibility in obsessive compulsive disorder using multilevel meta-analysis. *Clinical Psychological Science*, *6*, 481–505.

Frensch, P. A., & Buchner, A. (1999). Domain-generality versus domain-specificity in cognition. *The Nature of Cognition*, 137–172.

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

Galimberti, E., Martoni, R. M., Cavallini, M. C., Erzegovesi, S., & Bellodi, L. (2012). Motor inhibition and cognitive flexibility in eating disorder subtypes. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *36*, 307–312.

Gallagher, D., Parker, A., Samavat, H., & Zelig, R. (2022). Prophylactic supplementation of phosphate, magnesium, and potassium for the prevention of refeeding syndrome in hospitalized individuals with anorexia nervosa. *Nutrition in Clinical Practice*, *37*, 328–343.

Galmiche, M., Déchelotte, P., Lambert, G., & Tavolacci, M. P. (2019). Prevalence of eating disorders over the 2000–2018 period: A systematic literature review. *The American Journal of Clinical Nutrition*, *109*, 1402–1413.

Garcia-Burgos, D. (2022). The role of associative learning in current paradigm shifts in eating disorder research and clinical practice. MDPI.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating

Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*, 871–878.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31.

Geisler, D., Ritschel, F., King, J. A., Bernardoni, F., Seidel, M., Boehm, I., et al.others. (2017). Increased anterior cingulate cortex response precedes behavioural adaptation in anorexia nervosa. *Scientific Reports*, *7*, 42066.

Giel, K. E., Friederich, H.-C., Teufel, M., Hautzinger, M., Enck, P., & Zipfel, S. (2011). Attentional processing of food pictures in individuals with anorexia nervosa—an eye-tracking study. *Biological Psychiatry*, *69*, 661–667.

Gillan, C. M., & Robbins, T. W. (2014). Goal-directed learning and obsessive–compulsive disorder. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130475.

Giommi, F., Bauer, P. R., Berkovich-Ohana, A., Barendregt, H., Brown, K. W., Gallagher, S., et al.others. (2023). The (in) flexible self: Psychopathology, mindfulness, and neuroscience. *International Journal of Clinical and Health Psychology*, *23*, 100381.

Glashouwer, K. A., Bloot, L., Veenstra, E. M., Franken, I. H., & Jong, P. J. de. (2014). Heightened sensitivity to punishment and reward in anorexia nervosa. *Appetite*, *75*, 97–102.

Grant, D. A., & Berg, E. A. (1993). Wisconsin card sorting test. *Journal of Experimental Psychology*.

Grau, A., Magallón-Neri, E., Faus, G., & Feixas, G. (2019). Cognitive impairment in eating disorder patients of short and long-term duration: A case–control study. *Neuropsychiatric Disease and Treatment*, 1329–1341.

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*, 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.

Hagan, K. E., & Forbush, K. T. (2021). Reward learning in unmedicated women with bulimia nervosa: A pilot investigation. *Journal of Psychiatric Research*, *136*, 63–70.

Harrison, A., Genders, R., Davies, H., Treasure, J., & Tchanturia, K. (2011). Experimental measurement of the regulation of anger and aggression in women with anorexia nervosa. *Clinical Psychology & Psychotherapy*, *18*, 445–452.

Harvey, A. G., Watkins, E., & Mansell, W. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford University Press, USA.

Hatch, A., Madden, S., Kohn, M. R., Clarke, S., Touyz, S., Gordon, E., & Williams, L. M. (2010). In first presentation adolescent anorexia nervosa, do cognitive markers of underweight status change with weight gain following a refeeding intervention? *International Journal of Eating Disorders*, *43*, 295–306.

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. (2022). Beyond description and deficits: How computational psychiatry can enhance an understanding of decision-making in anorexia nervosa. *Current Psychiatry Reports*, 1–11.

Hekmat, S., Mehrjerdi, Z. A., Moradi, A., Ekhtiari, H., & Bakhshi, S. (2011). Cognitive flexibility, attention and speed of mental processing in opioid and methamphetamine addicts in comparison with non-addicts. *Basic and Clinical Neuroscience*, *2*, 12.

Hildebrandt, T., Grotzinger, A., Reddan, M., Greif, R., Levy, I., Goodman, W., & Schiller, D. (2015). Testing the disgust conditioning theory of food-avoidance in adolescents with recent onset anorexia nervosa. *Behaviour Research and Therapy*, *71*, 131–138.

Hildebrandt, T., Schulz, K., Schiller, D., Heywood, A., Goodman, W., & Sysko, R. (2018). Evidence of prefrontal hyperactivation to food-cue reversal learning in adolescents with anorexia nervosa. *Behaviour Research and Therapy*, *111*, 36–43.

Hirst, R. B., Beard, C. L., Colby, K. A., Quittner, Z., Mills, B. M., & Lavender, J.

M. (2017). Anorexia nervosa and bulimia nervosa: A meta-analysis of executive functioning. *Neuroscience & Biobehavioral Reviews*, *83*, 678–690.

Hitchcock, P. F., Fried, E. I., & Frank, M. J. (2022). Computational psychiatry needs time and context. *Annual Review of Psychology*, *73*, 243–270.

Jappe, L. M., Frank, G. K., Shott, M. E., Rollin, M. D., Pryor, T., Hagman, J. O., … Davis, E. (2011). Heightened sensitivity to reward and punishment in anorexia nervosa. *International Journal of Eating Disorders*, *44*, 317–324.

Jazbec, S., Pantelis, C., Robbins, T., Weickert, T., Weinberger, D. R., & Goldberg, T. E. (2007). Intra-dimensional/extra-dimensional set-shifting performance in schizophrenia: Impact of distractors. *Schizophrenia Research*, *89*, 339–349.

Jonker, N. C., Glashouwer, K. A., Hoekzema, A., Ostafin, B. D., & Jong, P. J. de. (2019). Attentional engagement with and disengagement from food cues in anorexia nervosa. *Behaviour Research and Therapy*, *114*, 15–24.

Jonker, N. C., Glashouwer, K. A., & Jong, P. J. de. (2022). Punishment sensitivity and the persistence of anorexia nervosa: High punishment sensitivity is related to a less favorable course of anorexia nervosa. *International Journal of Eating Disorders*, *55*, 697–702.

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*, 865–878.

Kass, A. E., Kolko, R. P., & Wilfley, D. E. (2013). Psychological treatments for eating disorders. *Current Opinion in Psychiatry*, *26*, 549–555.

Keating, C. (2010). Theoretical perspective on anorexia nervosa: The conflict of reward. *Neuroscience & Biobehavioral Reviews*, *34*, 73–79.

Keating, C., Tilbrook, A. J., Rossell, S. L., Enticott, P. G., & Fitzgerald, P. B. (2012). Reward processing in anorexia nervosa. *Neuropsychologia*, *50*, 567–575.

Kittel, R., Brauhardt, A., & Hilbert, A. (2015). Cognitive and emotional functioning in binge-eating disorder: A systematic review. *International Journal of Eating Disorders*, *48*, 535–554.

Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH research domain criteria initiative:

Background, issues, and pragmatics. *Psychophysiology*, *53*, 286–297.

Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*, 155–177.

Lang, P., & Bradley, M. M. (2007). The international affective picture system (IAPS) in the study of emotion and attention. *Handbook of Emotion Elicitation and Assessment*, *29*, 70–73.

Lee, J. E., Namkoong, K., & Jung, Y.-C. (2017). Impaired prefrontal cognitive control over interference by food images in binge-eating disorder and bulimia nervosa. *Neuroscience Letters*, *651*, 95–101.

Linardon, J., Fairburn, C. G., Fitzsimmons-Craft, E. E., Wilfley, D. E., & Brennan, L. (2017). The empirical status of the third-wave behaviour therapies for the treatment of eating disorders: A systematic review. *Clinical Psychology Review*, *58*, 125–140.

Loeber, S., Grosshans, M., Korucuoglu, O., Vollmert, C., Vollstädt-Klein, S., Schneider, S., … Kiefer, F. (2012). Impairment of inhibitory control in response to food-associated cues and attentional bias of obese participants and normal-weight controls. *International Journal of Obesity*, *36*, 1334–1339.

Lombardo, C. (2008). Adattamento italiano della multidimensional perfectionism scale (MPS). *Psicoterapia Cognitiva e Comportamentale*, *14*, 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*, 335–343.

Matera, C., Nerini, A., & Stefanile, C. (2013). The role of peer influence on girls' body dissatisfaction and dieting. *Revue Européenne De Psychologie Appliquée/European Review of Applied Psychology*, *63*, 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*, 455–470.

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

McAnarney, E. R., Zarcone, J., Singh, P., Michels, J., Welsh, S., Litteer, T., … Klein, J. D. (2011). Restrictive anorexia nervosa and set-shifting in adolescents: A biobehavioral interface. *Journal of Adolescent Health*, *49*, 99–101.

Mehler, P. S., & Brown, C. (2015). Anorexia nervosa–medical complications. *Journal of Eating Disorders*, *3*, 1–8.

Michelini, G., Palumbo, I. M., DeYoung, C. G., Latzman, R. D., & Kotov, R. (2021). Linking RDoC and HiTOP: A new interface for advancing psychiatric nosology and neuroscience. *Clinical Psychology Review*, *86*, 102025.

Miles, S., Gnatt, I., Phillipou, A., & Nedeljkovic, M. (2020). Cognitive flexibility in acute anorexia nervosa and after recovery: A systematic review. *Clinical Psychology Review*, *81*, 101905.

Miles, S., Howlett, C. A., Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the wisconsin card sorting test to assess cognitive flexibility. *Behavior Research Methods*, *53*, 2083–2091.

Miranda-Olivos, R., Testa, G., Lucas, I., Sanchez, I., Sanchez-Gonzalez, J., Granero, R., … Fernandez-Aranda, F. (2021). Clinical factors predicting impaired executive functions in eating disorders: The role of illness duration. *Journal of Psychiatric Research*, *144*, 87–95.

Mobbs, O., Iglesias, K., Golay, A., & Van der Linden, M. (2011). Cognitive deficits in obese persons with and without binge eating disorder. Investigation using a mental flexibility task. *Appetite*, *57*, 263–271.

Monni, A., Scandola, M., Hélie, S., & Scalas, L. F. (2023). Cognitive flexibility assessment with a new reversal learning task paradigm compared with the wisconsin card sorting test exploring the moderating effect of gender and stress. *Psychological Research*, *87*, 1439–1453.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140.

Monteleone, A. M., Monteleone, P., Esposito, F., Prinster, A., Volpe, U., Cantone, E., et al.others. (2017). Altered processing of rewarding and aversive basic taste stimuli in symptomatic women with anorexia nervosa and bulimia nervosa: An fMRI study. *Journal of Psychiatric Research*, *90*, 94–101.

Mora-Maltas, B., Lucas, I., Granero, R., Vintró-Alcaraz, C., Miranda-Olivos, R., Baenas, I., et al.others. (2023). Cognitive flexibility and DSM-5 severity criteria for eating disorders: Assessing drive for thinness and duration of illness as alternative severity variables. *Journal of Eating Disorders*, *11*, 155.

Morris, L., & Mansell, W. (2018). A systematic review of the relationship between rigidity/flexibility and transdiagnostic cognitive and behavioral processes that maintain psychopathology. *Journal of Experimental Psychopathology*, *9*, 2043808718779431.

Mukherjee, D., Filipowicz, A. L., Vo, K., Satterthwaite, T. D., & Kable, J. W. (2020). Reward and punishment reversal-learning in major depressive disorder. *Journal of Abnormal Psychology*, *129*, 810.

Nyhus, E., & Barceló, F. (2009). The wisconsin card sorting test and the cognitive assessment of prefrontal executive functions: A critical update. *Brain and Cognition*, *71*, 437–451.

Orellana, G., & Slachevsky, A. (2013). Executive functioning in schizophrenia. *Frontiers in Psychiatry*, *4*, 35.

Park, J., & Moghaddam, B. (2017). Impact of anxiety on prefrontal cortex encoding of cognitive flexibility. *Neuroscience*, *345*, 193–202.

Pearson, C. M., Wonderlich, S. A., & Smith, G. T. (2015). A risk and maintenance model for bulimia nervosa: From impulsive action to compulsive behavior. *Psychological Review*, *122*, 516.

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.

Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task? *Infant and Child Development: An International Journal of Research and Practice*, *11*, 93–105.

Perpiñá, C., Segura, M., & Sánchez-Reales, S. (2017). Cognitive flexibility and decision-making in eating disorders and obesity. *Eating and Weight Disorders*, *22*, 435–444.

Popescu, M., Popescu, E.-A., DeGraba, T. J., & Hughes, J. D. (2023). Cognitive flexibility in post-traumatic stress disorder: Sustained interference associated with altered modulation of cortical oscillatory activity during task-switching. *NeuroImage: Clinical*, *37*, 103297.

Pringle, A., Harmer, C., & Cooper, M. (2010). Investigating vulnerability to eating disorders: Biases in emotional processing. *Psychological Medicine*, *40*, 645–655.

Qian, J., Wu, Y., Liu, F., Zhu, Y., Jin, H., Zhang, H., ... Yu, D. (2022). An update on the prevalence of eating disorders in the general population: A systematic review and meta-analysis. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *27*, 415–428.

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

Raven, J. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment* (pp. 223–237). Springer.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–69). New York, NY: Appleton-Century Crofts.

Ritschel, F., Geisler, D., King, J. A., Bernardoni, F., Seidel, M., Boehm, I., ... Ehrlich, S. (2017). Neural correlates of altered feedback learning in women recovered from anorexia nervosa. *Scientific Reports*, *7*, 1–10.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., & RABBITT, P. M. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the*

International Neuropsychological Society, *4*, 474–490.

Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L., & Treasure, J. (2007). A systematic review and meta-analysis of set-shifting ability in eating disorders. *Psychological Medicine*, *37*, 1075–1084.

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

Sarrar, L., Holzhausen, M., Warschburger, P., Pfeiffer, E., Lehmkuhl, U., & Schneider, N. (2016). Cognitive function in adolescent patients with anorexia nervosa and unipolar affective disorders. *European Eating Disorders Review*, *24*, 232–240.

Schaefer, L. M., & Steinglass, J. E. (2021). Reward learning through the lens of RDoC: A review of theory, assessment, and empirical findings in the eating disorders. *Current Psychiatry Reports*, *23*, 1–11.

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.

Selby, E. A., & Coniglio, K. A. (2020). Positive emotion and motivational dynamics in anorexia nervosa: A positive emotion amplification model (PE-AMP). *Psychological Review*, *127*, 853–890.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772.

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.

Smink, F. R., Hoeken, D. van, & Hoek, H. W. (2013). Epidemiology, course, and outcome of eating disorders. *Current Opinion in Psychiatry*, *26*, 543–548.

Smith, K. E., Mason, T. B., Johnson, J. S., Lavender, J. M., & Wonderlich, S. A. (2018). A systematic review of reviews of neurocognitive functioning in eating disorders: The state-of-the-literature and future directions. *International Journal of Eating Disorders*, *51*, 798–821.

Spitoni, G. F., Aragonaa, M., Bevacqua, S., Cotugno, A., & Antonucci, G. (2018). An ecological approach to the behavioral assessment of executive functions in anorexia nervosa. *Psychiatry Research*, *259*, 283–288.

Steding, J., Boehm, I., King, J. A., Geisler, D., Ritschel, F., Seidel, M., et al.others. (2019). Goal-directed vs. Habitual instrumental behavior during reward processing in anorexia nervosa: An fMRI study. *Scientific Reports*, *9*, 13529.

Steinke, A., Lange, F., & Kopp, B. (2020). Parallel model-based and model-free reinforcement learning for card sorting performance. *Scientific Reports*, *10*, 15464.

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

Sunday, S. R., Halmi, K. A., & Einhorn, A. (1995). The yale-brown-cornell eating disorder scale: A new scale to assess eating disorder symptomatology. *International Journal of Eating Disorders*, *18*, 237–245.

Tchanturia, K., Campbell, I. C., Morris, R., & Treasure, J. (2005). Neuropsychological studies in anorexia nervosa. *International Journal of Eating Disorders*, *37*, S72–S76.

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

Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., ... Morris, R. (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PloS One*, *7*, e28331.

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*, 227–236.

Tchanturia, K., Harrison, A., Davies, H., Roberts, M., Oldershaw, A., Nakazato, M., ... Treasure, J. (2011). Cognitive flexibility and clinical severity in eating disorders. *Plos One*, *6*, e20462.

Tchanturia, K., Lounes, N., & Holttum, S. (2014). Cognitive remediation in anorexia nervosa and related conditions: A systematic review. *European Eating Disorders Review*, *22*, 454–462.

Tenconi, E., Santonastaso, P., Degortes, D., Bosello, R., Titton, F., Mapelli, D., & Favaro, A. (2010). Set-shifting abilities, central coherence, and handedness in anorexia nervosa patients, their unaffected siblings and healthy controls: Exploring putative endophenotypes. *The World Journal of Biological Psychiatry*, *11*, 813–823.

Teng, E. J., Woods, D. W., & Twohig, M. P. (2006). Habit reversal as a treatment for chronic skin picking: A pilot investigation. *Behavior Modification*, *30*, 411–422.

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*, 683.

Van Den Berg, E., Nys, G. M., Brands, A. M., Ruis, C., Van Zandvoort, M. J., & Kessels, R. P. (2009). The brixton spatial anticipation test as a test for executive function: Validity in patient groups and norms for older adults. *Journal of the International Neuropsychological Society*, *15*, 695–703.

Verdejo-Garcia, A., Clark, L., Verdejo-Roman, J., Albein-Urios, N., Martinez-Gonzalez, J. M., Gutierrez, B., & Soriano-Mas, C. (2015). Neural substrates of cognitive flexibility in cocaine and gambling addictions. *The British Journal of Psychiatry*, *207*, 158–164.

Visu-Petra, L., & Mărcuș, O. (2019). Cognitive (in) flexibility: Potential risk factor for internalizing psychopathology across development. *P., Buchwald, KA, Moore, K., Kaniasty, P. Arenas-Landgrave,(Eds.), Stress and Anxiety: Contributions of the STAR Award Winners*, 125–134.

Wagner, A., Aizenstein, H., Venkatraman, V. K., Fudge, J., May, J. C., Mazurkewicz, L., et al.others. (2007). Altered reward processing in women recovered from anorexia nervosa. *American Journal of Psychiatry*, *164*, 1842–1849.

Walsh, B. T. (2013). The enigmatic persistence of anorexia nervosa. *American Journal of Psychiatry*, *170*, 477–484.

Waltz, J. A. (2017). The neural underpinnings of cognitive flexibility and their disruption in psychotic illness. *Neuroscience*, *345*, 203–217.

Watkins, E. (2015). An Alternative Transdiagnostic Mechanistic Approach to Affective Disorders Illustrated With Research From Clinical Psychology. *Emotion Review*, *7*, 250–255.

Weiss, E. O., Kruppa, J. A., Fink, G. R., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2021). Developmental differences in probabilistic reversal learning: A computational modeling approach. *Frontiers in Neuroscience*, *14*, 536596.

Wierenga, C. E., Ely, A., Bischoff-Grethe, A., Bailer, U. F., Simmons, A. N., & Kaye, W. H. (2014). Are extremes of consumption in eating disorders related to an altered balance between reward and inhibition? *Frontiers in Behavioral Neuroscience*, *8*, 410.

Wildes, J. E., Forbes, E. E., & Marcus, M. D. (2014). Advancing research on cognitive flexibility in eating disorders: The importance of distinguishing attentional set-shifting and reversal learning. *International Journal of Eating Disorders*, *47*, 227–230.

Wollenhaupt, C., Wilke, L., Erim, Y., Rauh, M., Steins-Loeber, S., & Paslakis, G. (2019). The association of leptin secretion with cognitive performance in patients with eating disorders. *Psychiatry Research*, *276*, 269–277.

Wonderlich, J. A., Bershad, M., & Steinglass, J. E. (2021). Exploring neural mechanisms related to cognitive control, reward, and affect in eating disorders: A narrative review of FMRI studies. *Neuropsychiatric Disease and Treatment*, 2053–2062.

World Health Organization. (2018). *International classification of diseases 11th revision (ICD-11)*. World Health Organization.

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*, 3365–3385.

Zhang, Z., Manson, K. F., Schiller, D., & Levy, I. (2014). Impaired associative learning with food rewards in obese women. *Current Biology*, *24*, 1731–1736.

Appendix A

DIC of the Hierarchical Reinforcement Learning Drift Diffusion Model

The HDDMrl, utilizing established literature priors and the Markov Chain Monte Carlo method, was fit to the data to estimate the joint posterior distribution of all parameters. The goal was to explore the interference of disorder-related information on decision-making by comparing various HDDMrls, which conditioned none, some, or all model parameters on diagnostic group and image categories. The Deviance Information Criterion (DIC) of each model was computed to identify the model offering the optimal balance between fit quality and complexity, (*i.e.*, the model with the lowest DIC).

The models explored include:

- M1: A standard HDDMrl.
- M2: Builds on M1 by having separate learning rates for positive and negative reinforcements.
- M3: The parameters α^+ and α^- are based on the diagnostic group.
- M4: Expands M3 by conditioning α^+ and α^- on both diagnostic group and image category.
- M5: Extends M4, considering the a parameter's potential influence by both diagnostic group and image category.
- M6: Builds on M5 by evaluating the v parameter's possible influence by both categories.
- M7: Extends M6 by considering the t parameter's dependence on both categories.

- M8: Adds to M7 by estimating a potential bias in the z parameter.

Model M7 emerged as the winning HDDMrl with the lowest DIC, where parameters α^+, α^-, a, v, t (excluding z) are conditioned on both diagnostic group and image category, offering insights into how disease-related information might affect decision processes.

Model	DIC
M1	103209.264
M2	101590.157
M3	101613.877
M4	99133.675
M5	96150.581
M6	95434.070
M7	**92808.856**
M8	93157.611

Correlation Matrices Between HDDMrl Learning Rate Parameters (α^- and α^+) and Key Clinical Measures in Eating Disorder Studies

To elucidate the role of reversal learning deficits, particularly in the context of eating disorders, a comprehensive analysis was conducted using correlation matrices. These matrices linked psychological measures in eating disorders with key parameters from the HDDMrl, namely α^- and α^+, which signify learning rates from punished and rewarded PRL-trials, respectively. Distinct matrices were constructed for neutral and food choices, each providing insights into the relationship between clinical aspects of eating disorders and these two learning rates.

Focusing on food choices, the correlation matrix revealed interesting patterns. The negative learning rate α^- displays a generally negative correlation with clinical measures, including a notable $r =$ -0.45 with the Eating Attitude Test-26 total score ($EAT-26_{tot}$) and its subscales - Dieting ($EAT-26_d$), Bulimia ($EAT-26_b$), and Oral Control ($EAT-26_{oc}$). This finding suggests that individuals with higher levels of disordered eating behaviors, as indicated by higher scores on the EAT-26, tend to have a decreased learning rate when faced with negative outcomes or punishments, particularly in contexts involving food-related choices. Conversely, individuals with lower EAT-26 scores, indicating fewer disordered eating behaviors, are likely to exhibit a higher learning rate in response to punitive actions.

In a parallel vein, the positive learning rate (α^+) also presents negative correlations with these clinical measures. For instance, its correlation with the EAT-26 total score and the Oral Control subscale, both quantified at $r = -0.45$, suggests that, similar to situations involving punishment, individuals with eating disorders exhibit a lower learning rate when rewards are presented in scenarios involving food choices.

This implies that healthier individuals tend to perform more effectively in PRL tasks, regardless of whether the outcome is a reward or a punishment, in the presence of food choices. This pattern may mirror how reward-driven behaviors influence eating patterns and attitudes towards food, offering insight into the complex interplay between psychological responses and dietary habits.

In contrast, the matrix focusing on neutral choices portrayed a different pattern. The negative learning rate α^- generally showed weak negative correlations with clinical measures, suggesting a subtle link with factors related to eating behaviors and body image perceptions under neutral conditions. Conversely, the positive learning rate α^+ displayed mostly minimal correlations, indicating a very limited impact of these psychological measures on learning processes in neutral choice scenarios. Compared to the matrix for food choices, this suggests that the influence of learning rates on eating disorder-related psychological measures might be domain-specific, with neutral choices eliciting a different pattern of associations.

Additionally, the study found that the correlations between both negative (α^-) and positive (α^+) learning rates, and other measures related to eating disorders were generally weak or even negligible, in both neutral and food-related scenarios. These measures include the Body Shape Questionnaire-14 ($BSQ - 14$) and the subscales of the Multidimensional Perfectionism Scale (MPS_{tot}), namely Concerns over Mistakes and Doubts (MPS_{cmd}), Personal Standards (MPS_{ps}), Parental Expectations and Criticism (MPS_{pec}), and Organization (MPS_{or}).

Overall, the correlation analysis reveals that (1) the most important clinical dimensions, closely linked to learning processes, as indicated by α^- and α^+ scores, are related to dysfunctional eating habits ($EAT - 26$). These show a stronger correlation than concerns about body image and perfectionism, as measured respectively by the $BSQ - 14$ and MPS scales, where the correlations are negligible; (2) these matrices provide a nuanced understanding of the interactions between reinforced learning rates and psychological measures in the context of eating disorders. They highlight the complexity and domain-specific nature of these relationships, particularly given that strong results only emerge in food-related scenarios.

Table 2: Pearson Correlation Matrix for Food Choices in the entire sample. This matrix displays correlation coefficients between HDDMrl parameters $\alpha^{-} food$ and $\alpha^{+} food$ and key psychological measures, including EAT-26, BSQ-14, and MPS. Ranging from -1 to +1, these coefficients reveal the linear associations between learning rates and factors linked to eating behaviors and body image perceptions.

	$BSQ-14$	MPS_{cmd}	MPS_{ps}	MPS_{pec}	MPS_{or}	MPS_{tot}	$EAT-26_{tot}$	$EAT-26_{d}$	$EAT-26_{b}$	$EAT-26_{oc}$	α^{-}	α^{+}
$BSQ-14$	1.00											
MPS_{cmd}	0.60	1.00										
MPS_{ps}	0.24	0.42	1.00									
MPS_{pec}	0.37	0.43	0.18	1.00								
MPS_{or}	-0.05	0.01	0.27	-0.03	1.00							
MPS_{tot}	0.51	0.81	0.70	0.65	0.37	1.00						
$EAT-26_{tot}$	0.71	0.50	0.40	0.25	0.12	0.52	1.00					
$EAT-26_{d}$	0.81	0.53	0.33	0.27	0.04	0.49	0.92	1.00				
$EAT-26_{b}$	0.74	0.48	0.26	0.36	-0.04	0.45	0.87	0.84	1.00			
$EAT-26_{oc}$	0.34	0.34	0.42	0.09	0.23	0.41	0.81	0.57	0.56	1.00		
α^{-}	-0.24	-0.20	-0.27	-0.01	-0.01	-0.20	-0.45	-0.38	-0.34	-0.41	1.00	
α^{+}	-0.29	-0.13	-0.17	-0.01	0.04	-0.11	-0.45	-0.35	-0.30	-0.45	0.60	1.00

Table 3: Pearson Correlation Matrix for Neutral Choices in the entire sample. This matrix displays correlation coefficients between HDDMrl parameters $\alpha^{-}neutral$ and $\alpha^{+}neutral$ and key psychological measures, including EAT-26, BSQ-14, and MPS. Ranging from -1 to +1, these coefficients reveal the linear associations between learning rates and factors linked to eating behaviors and body image perceptions.

	$BSQ-14$	MPS_{cmd}	MPS_{ps}	MPS_{pec}	MPS_{or}	MPS_{tot}	$EAT-26_{tot}$	$EAT-26_{d}$	$EAT-26_{b}$	$EAT-26_{oc}$	α^{-}	α^{+}
$BSQ-14$	1.00											
MPS_{cmd}	0.58	1.00										
MPS_{ps}	0.22	0.41	1.00									
MPS_{pec}	0.36	0.40	0.15	1.00								
MPS_{or}	-0.08	0.00	0.30	-0.09	1.00							
MPS_{tot}	0.49	0.80	0.70	0.61	0.37	1.00						
$EAT-26_{tot}$	0.71	0.50	0.40	0.23	0.14	0.53	1.00					
$EAT-26_{d}$	0.80	0.53	0.32	0.26	0.07	0.50	0.92	1.00				
$EAT-26_{b}$	0.73	0.46	0.26	0.33	-0.05	0.43	0.87	0.83	1.00			
$EAT-26_{oc}$	0.36	0.36	0.43	0.09	0.25	0.44	0.82	0.58	0.57	1.00		
α^{-}	-0.12	0.04	-0.03	-0.15	0.04	-0.03	-0.21	-0.16	-0.21	-0.17	1.00	
α^{+}	-0.07	-0.04	-0.04	-0.06	0.09	-0.03	-0.08	-0.08	-0.04	-0.07	0.33	1.00

Appendix B

DIC of the Hierarchical Reinforcement Learning Drift Diffusion Model

In this analysis, the aim was to investigate the impact of symptom severity on decision-making by comparing different HDDMrls, each conditioning none, some, or all model parameters on the diagnostic group. The DIC was computed for each model to determine the one providing an optimal balance between fit quality and complexity, denoted by the model with the lowest DIC value.

The models explored include:

- M1: A standard HDDMrl.
- M2: Builds on M1 by having separate learning rates for positive and negative reinforcements.
- M3: The parameters α^+ and α^- are based on the diagnostic group.
- M4: Expands M3 by conditioning a on diagnostic group.
- M5: Extends M4, considering the v parameter's potential influence by diagnostic group.
- M6: Builds on M5 by evaluating the t parameter's possible influence by diagnostic group.
- M7: Adds to M6 by estimating a potential bias in the z parameter.

Model M6 emerged as the winning HDDMrl with the lowest DIC, where parameters α^+, α^-, a, v, t (excluding z) are conditioned on diagnostic group, offering insights into how symptom severity might affect decision processes.

Model	DIC
M1	27563.015849
M2	27027.093319
M3	27093.509238
M4	27208.442459
M5	27020.789676
M6	**27019.599486**
M7	27021.329592

Appendix C

DIC of the Hierarchical Reinforcement Learning Drift Diffusion Model

In this analysis, the aim was to investigate reversal learning processes in AN and HC group by comparing different HDDMrls, each conditioning none, some, or all model parameters on the diagnostic group. The DIC was computed for each model to determine the one providing an optimal balance between fit quality and complexity, denoted by the model with the lowest DIC value.

The models explored include:

- M1: A standard HDDMrl.
- M2: Builds on M1 by having separate learning rates for positive and negative reinforcements.
- M3: The parameters α^+ and α^- are based on the diagnostic group.
- M4: Expands M3 by conditioning a on diagnostic group.
- M5: Extends M4, considering the v parameter's potential influence by diagnostic group.
- M6: Builds on M5 by evaluating the t parameter's possible influence by diagnostic group.
- M7: Adds to M6 by estimating a potential bias in the z parameter.

Model M6 emerged as the winning HDDMrl with the lowest DIC, where parameters α^+, α^-, a, v, t (excluding z) are conditioned on diagnostic group.

Model	DIC
M1	15884.212575
M2	15226.061842
M3	15269.189945
M4	15332.841152
M5	15276.743301
M6	**15178.813533**
M7	15276.221193

DIC of the Drift Diffusion Model

In this analysis, the aim was to investigate the choice processes in AN and HC group by comparing different DDM, each conditioning none, some, or all model parameters on the diagnostic group. The DIC was computed for each model to determine the one providing an optimal balance between fit quality and complexity, denoted by the model with the lowest DIC value.

The models explored include:

- M1: A standard DDM.
- M2: Expands M1 by conditioning a on diagnostic group.
- M3: Extends M2, considering the v parameter's potential influence by diagnostic group.
- M4: Builds on M3 by evaluating the t parameter's possible influence by diagnostic group.
- M5: Builds on M4 by conditioning a on diagnostic group and switching effect (*switch* or *repeat*).
- M6: Builds on M5 by conditioning v on diagnostic group and switching effect.
- M7: Builds on M6 by conditioning t on diagnostic group and switching effect.
- M8: Adds to M7 by estimating a potential bias in the z parameter.

Model M7 emerged as the winning DDM with the lowest DIC, where parameters a, v, t (excluding z) are conditioned on diagnostic group.

Model	DIC
M1	45716.279934
M2	45718.713843
M3	45719.399631
M4	45718.466561
M5	34252.451721
M6	33524.920970
M7	**32783.616534**
M8	32791.887132

Appendix D

In evaluating the WCST performance using the wP-RL model, eight key parameters related to decision-making were analyzed. These parameters showed no credible difference between the AN-R patient group and the healthy controls, indicating comparable task performance by both groups (Figure 1; 2; 3; 4; 5; 6; 7; 8).

Moreover, a comparison between the proportions of perseverative errors and non-perseverative errors in AN-R patients and controls was conducted. These indices are the most commonly used in scoring the WCST in previous literature. No credible differences emerges between AN-R patients and HC, in the proportion of perseverative and non-perseverative errors (Figure 9; 10).

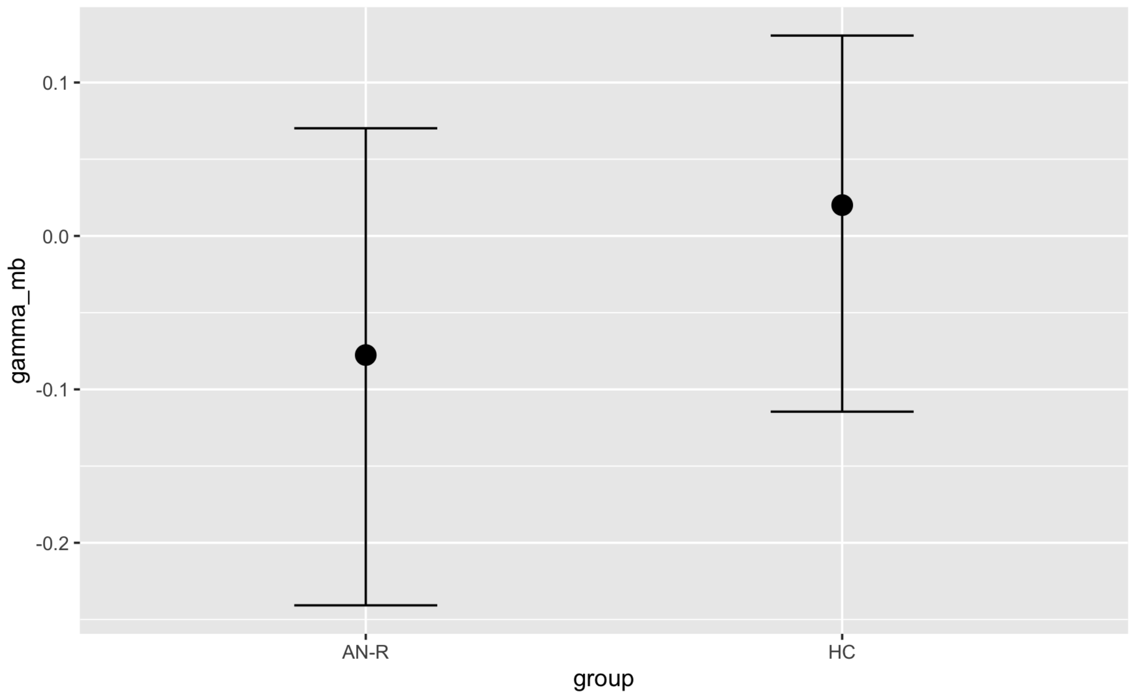

Figure 1: Mean of the posterior distribution for γ_{MB} parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

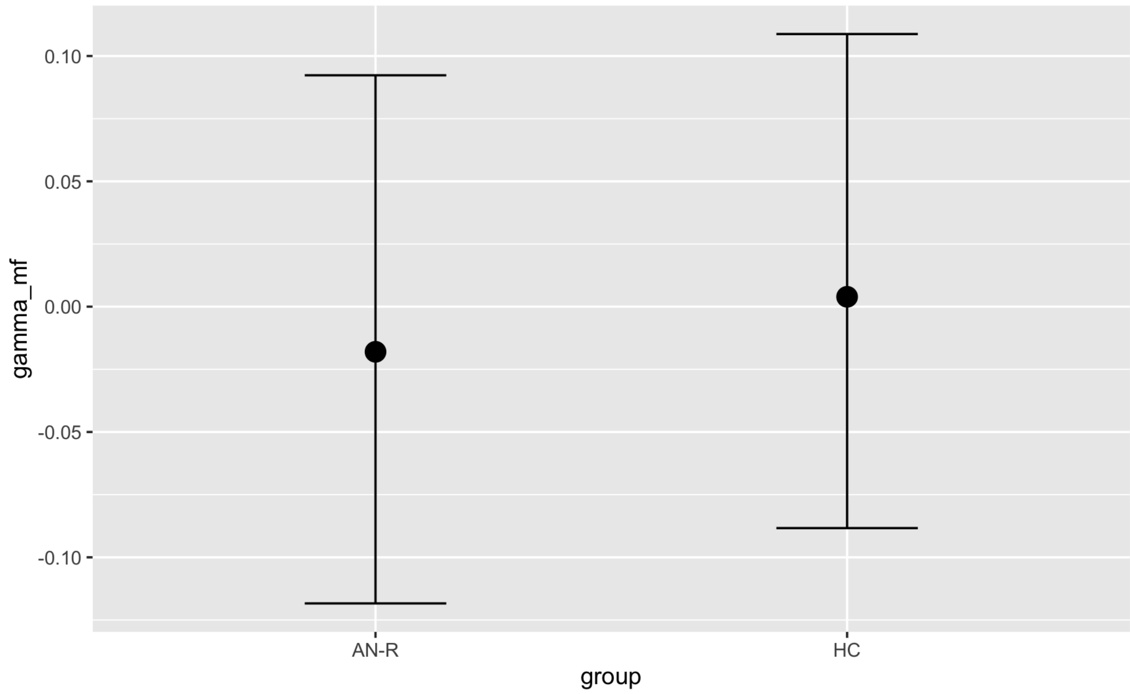

Figure 2: Mean of the posterior distribution for γ_{MF} parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

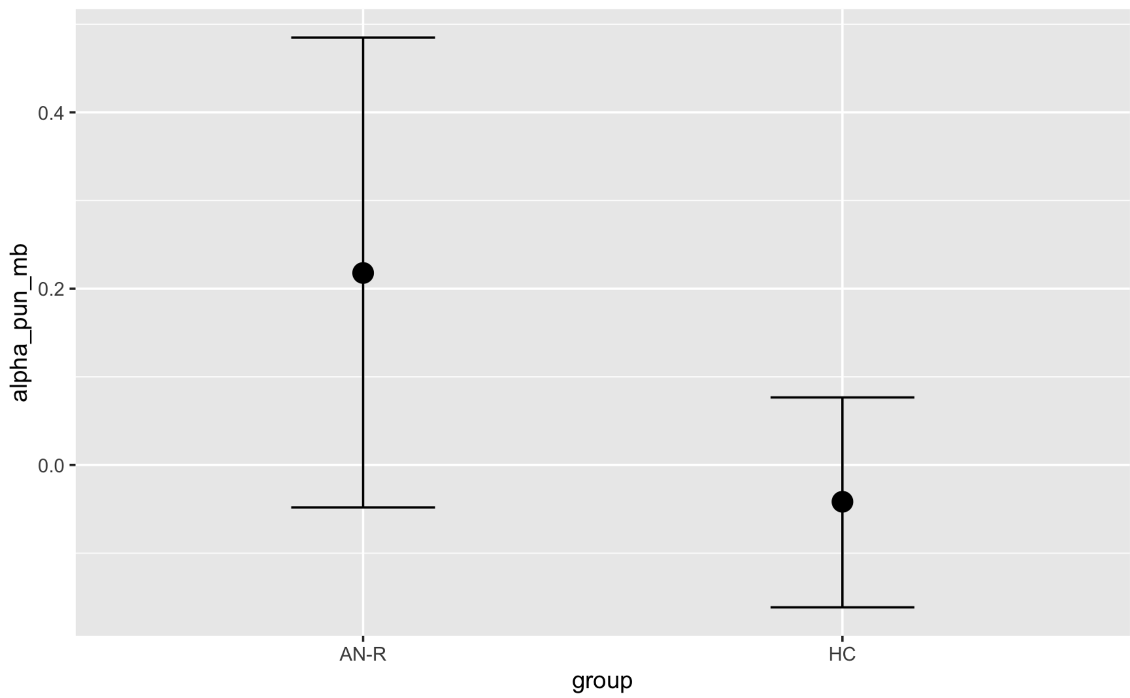

Figure 3: Mean of the posterior distribution for α^{-}_{MB} parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

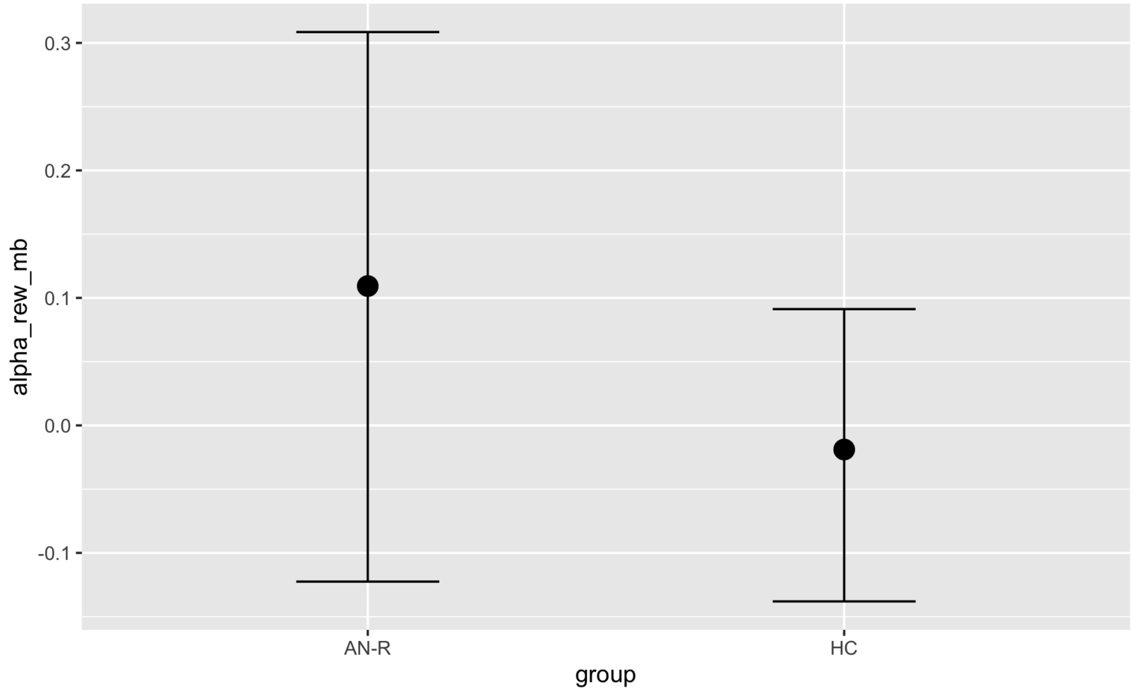

Figure 4: Mean of the posterior distribution for α^{+}_{MB} parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

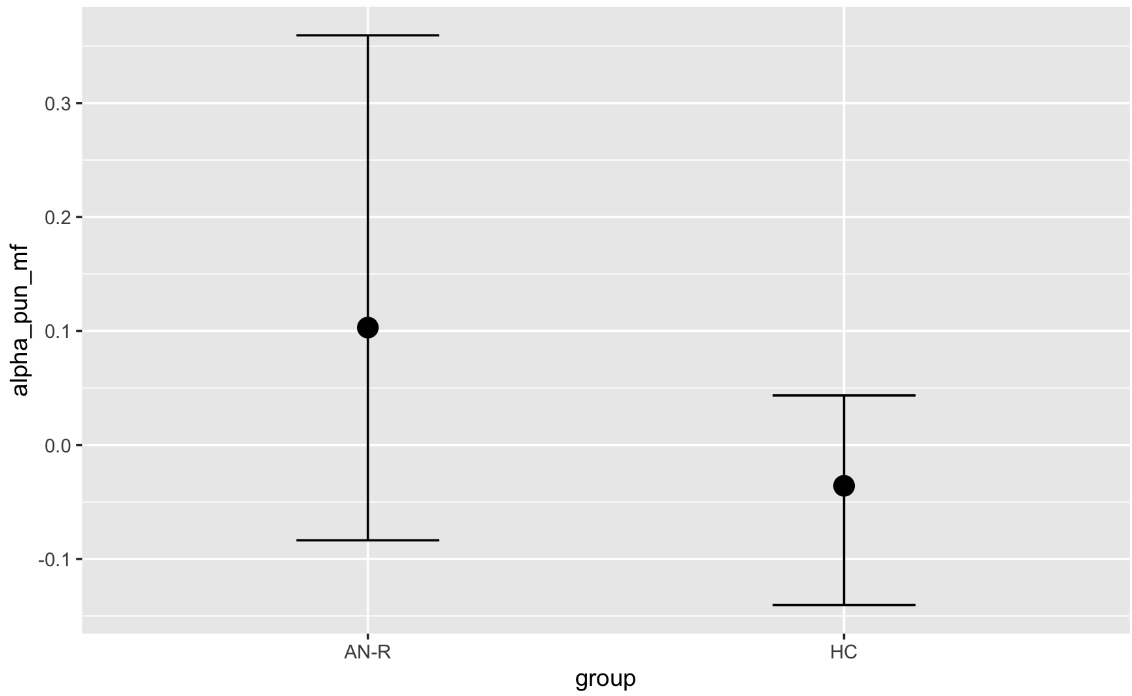

Figure 5: Mean of the posterior distribution for α^{-}_{MF} parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

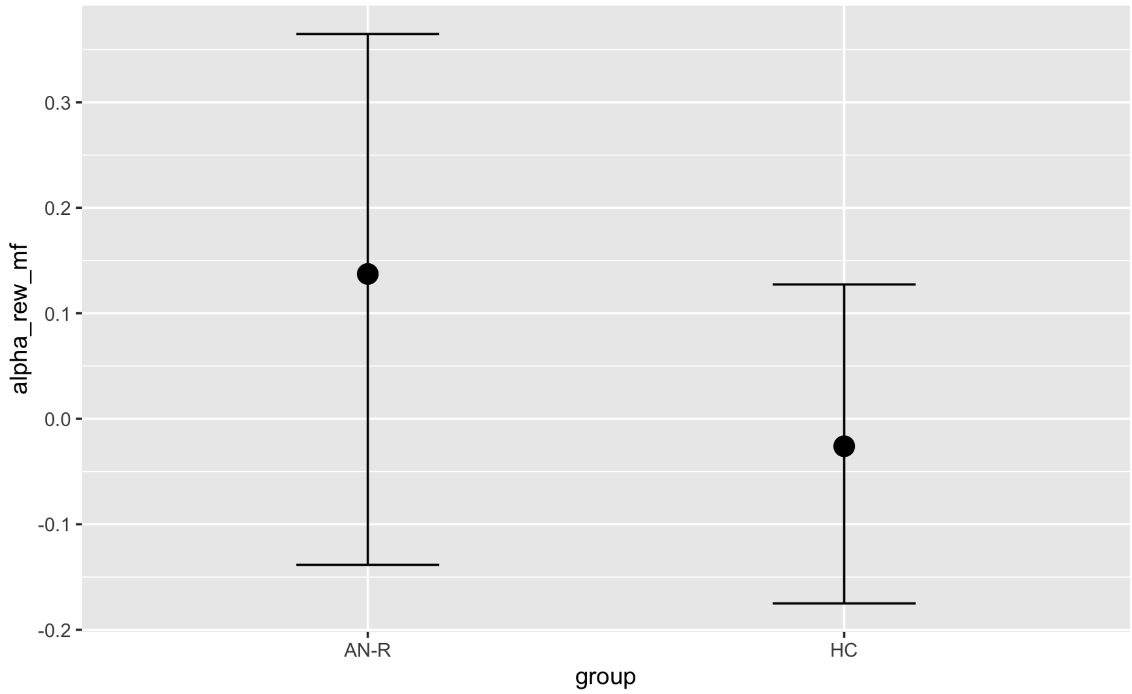

Figure 6: Mean of the posterior distribution for α^{+}_{MF} parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

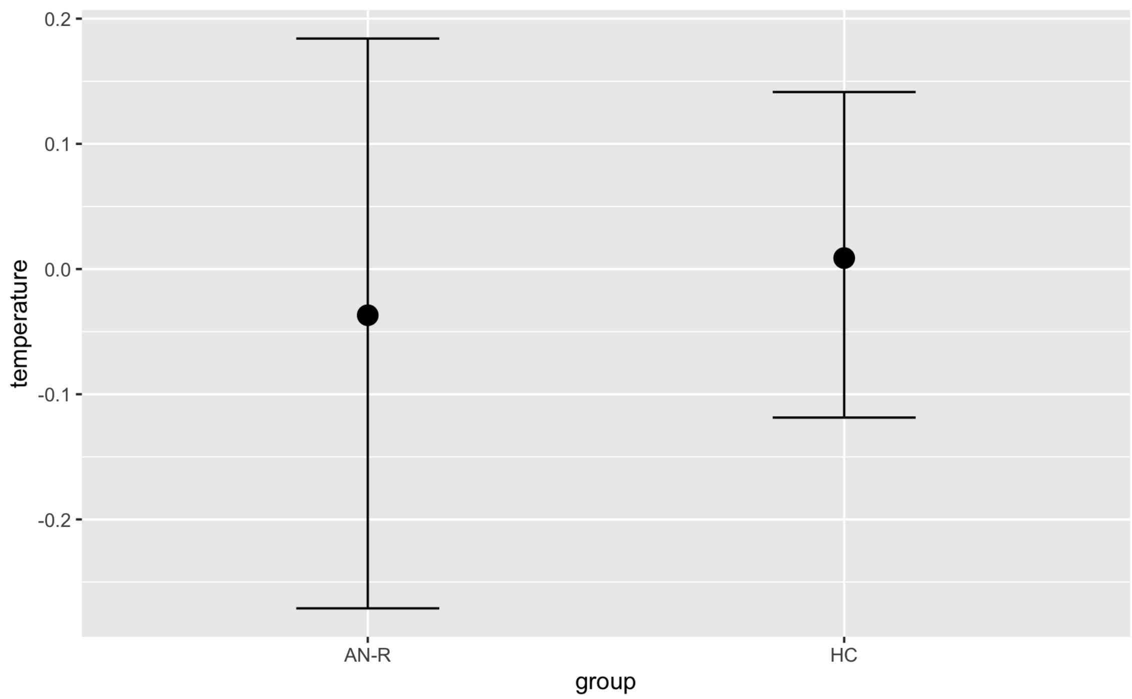

Figure 7: Mean of the posterior distribution for t parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

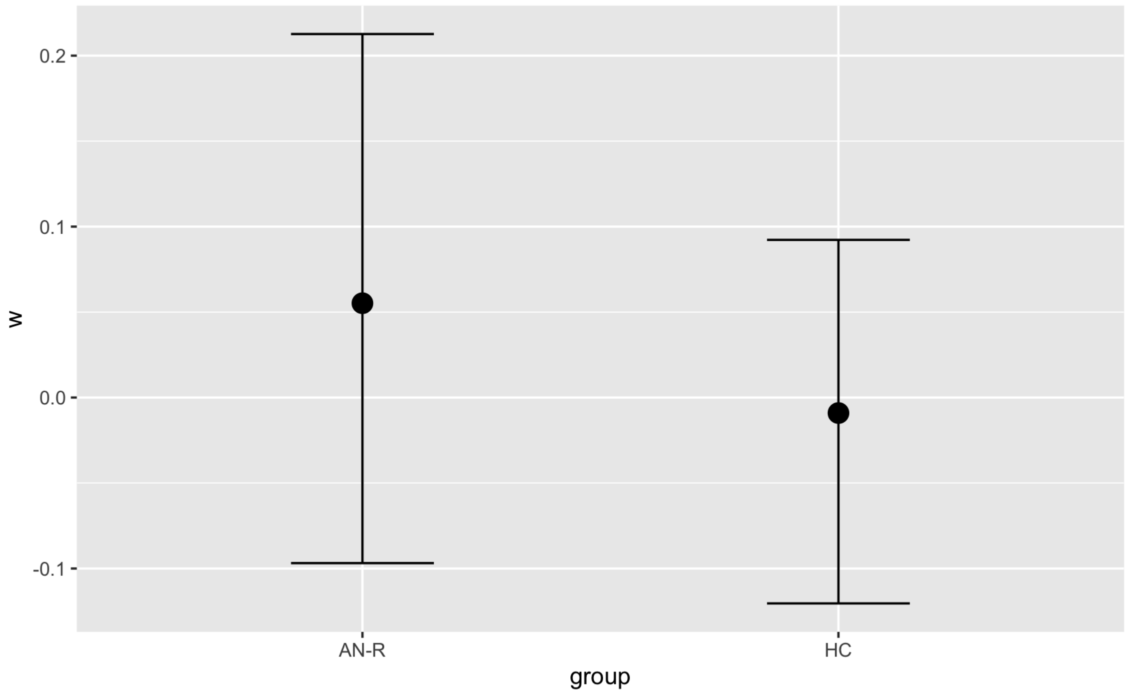

Figure 8: Mean of the posterior distribution for w parameter from the wP-RL model as a function of groups in the WCST. Vertical bars denote credible intervals.

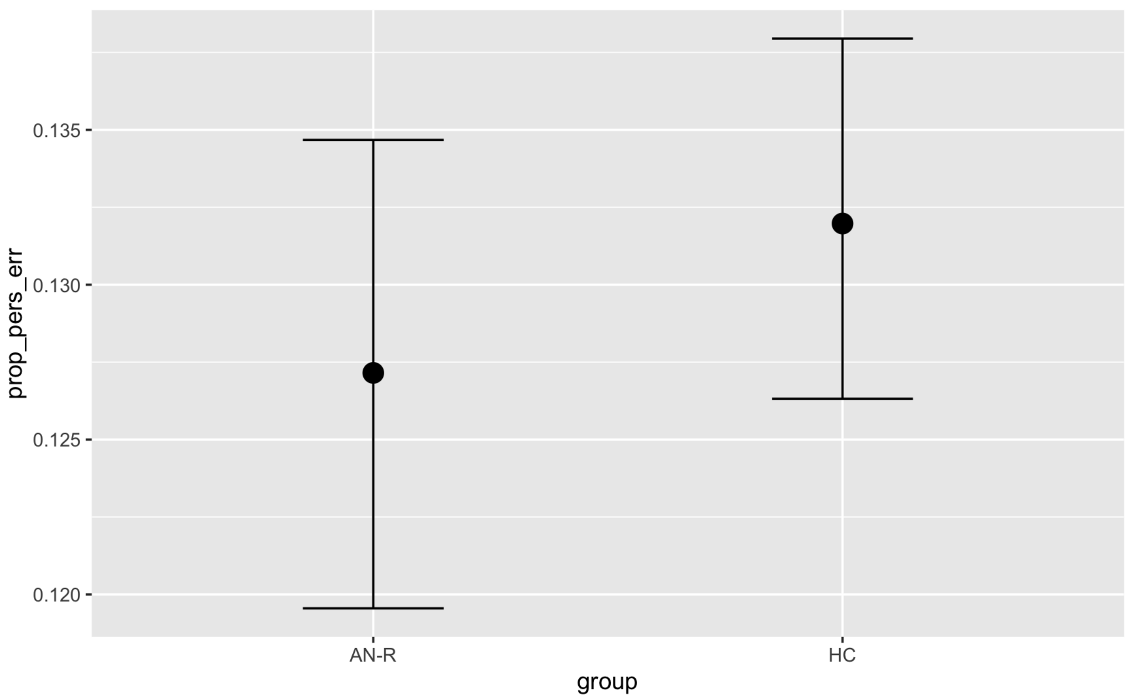

Figure 9: Mean of the posterior distribution for the proportion of *perseverative* errors as a function of groups in the WCST. Vertical bars denote credible intervals.

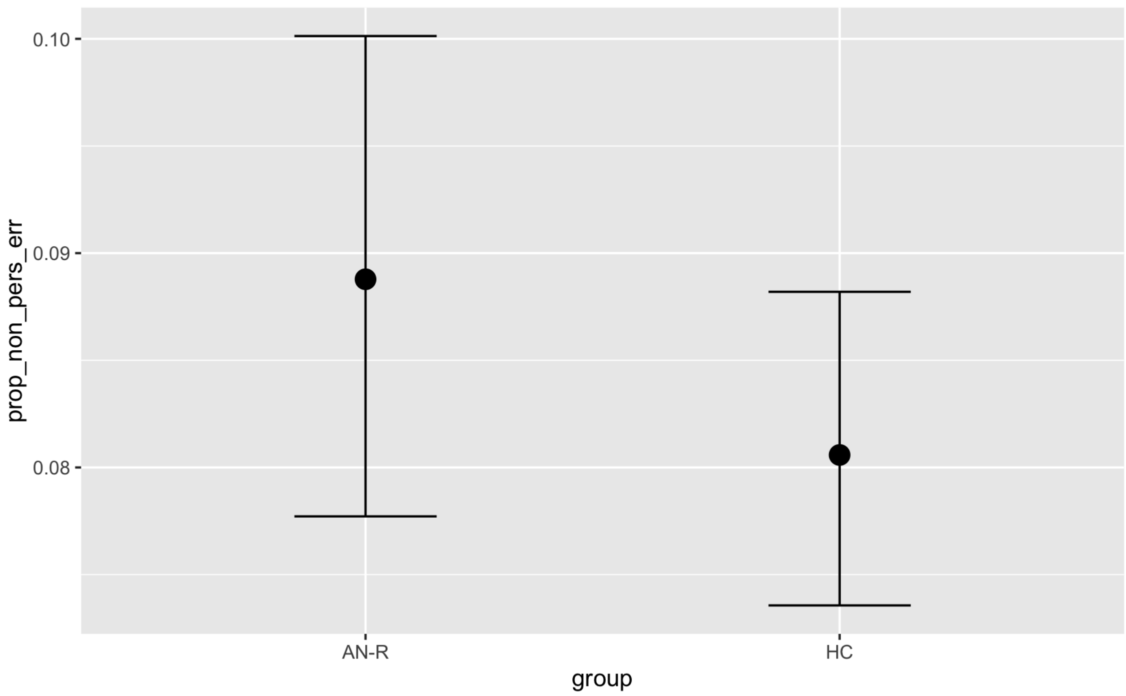

Figure 10: Mean of the posterior distribution for the proportion of $non-perseverative$ errors as a function of groups in the WCST. Vertical bars denote credible intervals.

Acknowledgments

I would like to express my profound gratitude to Prof. Claudio Sica and Prof. Corrado Caudek for their support and guidance in my Ph.D. journey.

I would also like to thank Dr. Lisa Guidi, Dr. Sara Danti, Dr. Silvia Lucchesi, Dr. Elisa Giusti, and Dr. Martina Di Meglio from Istituto Specchidacqua, Montecatini, Italy, for their valuable contributions to this study. My gratitude also extends to Dr. Caterina Pieraccioli from Neomesia, Villa dei Pini, Firenze, Italy.