A PILOT STUDY OF THE EFFECTS OF LISDEXAMFETAMINE DIMESYLATE (VYVANSE) TREATMENT ON REINFORCEMENT LEARNING IN ADULTS WITH BULIMIA NERVOSA

by

Victoria Brown

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2021

Dalhousie University is located in Mi'kma'ki,
the ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

© Copyright by Victoria Brown, 2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Although bulimia nervosa (BN) is thought to be associated with impairments in reward-related decision making, little is known about the decisional processes that contribute to BN behaviour. Twenty-three participants with moderate-extreme BN were administered a two-step reinforcement learning task to assess their relative degree of decisional exploration/exploitation, goal directed versus habitual control, and learning rate before and after treatment with lisdexamfetamine dimesylate (LDX). BN symptom changes were also monitored. Paired permutation tests show a statistically significant decrease in objective binge episodes (M_{T3}-M_{T1}= -35.11, $p < 0.001$) and compensatory behaviours (M_{T3}-M_{T1}= -38.85, $p< 0.001$) at maintenance drug dosage compared to baseline. However, reward learning, as far as it is assessed by the task did not seem to contribute to the effect of LDX on BN behaviour. It is hoped that these findings will contribute to an improved understanding of the computational nature of decision making among individuals with BN.

LIST OF ABBREVIATIONS/SYMBOLS USED

ADHD	Attention deficit hyperactivity disorder
AMPT	Alpha-methyl-para-tyrosine
ASRS	Adult ADHD Self-Report Scale
BDNF	Brain derived neurotrophic factor
BED	Binge eating disorder
BMI	Body mass index
BN	Bulimia nervosa
CB	Compensatory behaviours
COMT	Catechol-O-methyltransferase
COVID-19	Coronavirus disease
C-SSRS	Columbia Suicide Severity Rating Scale
DA	Dopamine
EDDS	Eating Disorder Diagnostic Scale
EDE 17.0D	Eating Disorder Examination 17.0D
EKG	Electrocardiogram
EL	Emotional lability
H_a	Alternative hypothesis
H_0	Null hypothesis
LDX	Lisdexamfetamine dimesylate
M	Sample mean
MAOI	Monoamine oxidase inhibitor
Met	Methionine
OBD	Objective binge days
OBE	Objective binge episodes
OSPAN	Operation Span Task
Q^{MB}	Model-based value function
Q^{MF}	Model-free value function
QTc	Corrected QT interval
RR	Reward rate
SARSA	State-action-reward-state-action
SBD	Subjective binge days
SBE	Subjective binge episodes
SCID-5-RV	Structured Clinical Interview for DSM-5 Disorders Research Version
SD	Standard deviation
SE	Standard error
T1	Time 1 (baseline)
T3	Time 3 (maintenance dose)
T4	Time 4 (post)
TEAE	Treatment emergent adverse events
WHO-CIDI	World Health Organization's Composite International Diagnostic Interview
WM	Working memory
α	Learning rate

β	Inverse softmax temperature
ρ	Perseveration
ω	Relative weighting of participant model-based and model-free strategies

ACKNOWLEDGEMENTS

This thesis would not have been possible without the kind support of several individuals. I would like to extend my sincerest gratitude to all of them.

Thank you to Dr. Abraham Nunes for all of your guidance and support throughout the writing of my thesis paper. I am so lucky to have been able to learn from you Abe.

Thank you to Dr. Keshen for your boundless encouragement throughout my thesis writing process. Thank you also for believing in me and affording me the opportunity of learning from you at the Eating Disorder Clinic. Given my own personal struggles with disordered eating, this was a dream come true.

Thank you to Dr. Gamberg for supervising my independent study project and for your endless support throughout my time as a Master's student. Your dedication and passion for your work is inspiring.

Thank you to my fellow researchers Laura Dixon, Sarrah Ali, and Emma Bodnar. We make a wonderful team. I would like to express my special gratitude to Laura, for her willingness to answer questions, and her time in helping me to navigate my Master's thesis requirements.

I would also like to express my gratitude toward Dr. Igor Yakovenko and Dr. Thomas Trappenberg for sharing their knowledge and providing feedback throughout the completion of this paper.

Finally, thank you to Alex, Mom, Dad, Pops, Grandma, and Grandpa and Grandma Brown. Thank you for loving me unconditionally and being there for me every step of the way. I owe every success to you.

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND

Bulimia nervosa (BN) is characterized by recurrent episodes of binge eating and inappropriate compensatory behaviours such as vomiting, laxative use, or excessive physical activity (Rushing et al., 2003). Although several empirically supported psychotherapies currently exist for BN patients (e.g., cognitive behavioural therapy), binge-purge abstinence is typically achieved in less than 50 percent of individuals receiving treatment (Wilson et al., 2007; Watson et al., 2018). One reason that these evidence-based treatments may be less effective for some is that they fail to address the underlying neurobiological mechanisms of BN (Hagan & Forbush, 2021). Presently, BN pathophysiology is poorly understood (Kekic et al., 2016), and therefore, is not adequately targeted by existing *gold-standard* therapies. Efforts to understand BN pathophysiology are warranted to develop effective treatments that target the neurobiological underpinnings of the disorder. The following sections explore the current literature on such underlying neurocognitive mechanisms, with a particular focus on reward-based decision making.

1.2 ELEMENTS OF REWARD-BASED DECISION-MAKING

Three putative reward-based decision-making deficits in BN involve the following neurocognitive factors: A) goal directed vs. habitual control balance, B) learning rate, and C) the exploration/exploitation trade-off.

1.2.1 Goal-Directed Versus Habitual Control

Evidence suggests that performance in reward-related behaviour is driven in part by two different processes, namely, goal-directed and habitual control (Balleine & O'Doherty, 2010). Goal-directed actions are governed by an expected outcome of a particular action, and the value assigned to that outcome (Furlong et al., 2014). This behaviour utilises model-based learning that acknowledges an internal model of cause and effect contingencies in the environment (Onysk & Series, 2020) and typically moves organisms toward goal-satisfaction (O'Doherty et al., 2017). Hypothetically, individuals with BN who exhibit greater goal-directed control would have increased awareness of the long-term consequences of their binge/purge behaviour (e.g. enamel erosion, esophageal ulcers, electrolyte changes, tachycardia, and impaired psychosocial function; Dynesen et al., 2008; Mehler & Rylander, 2015). With this heightened awareness of expected outcomes, individuals with BN might abstain from bulimia behaviors that lead to long term consequences, and instead, engage in alternative actions that could optimize positive outcomes (e.g., physical health and psychosocial functioning).

Unlike goal-directed behaviour, habitual actions are elicited by the perception of a stimulus (Thorndike, 1898), and occur even when the value of outcomes have lessened, and are therefore no longer rewarding (Steinglass & Walsh, 2016). Habitual behaviour

uses model-free learning which involves simply repeating actions that were previously rewarded, without requiring an internal model of cause and effect relationships in the environment. Though habitual responding is more cognitively efficient than goal-directed control (O'Doherty et al., 2017), the approach does not allow for refined decision making through more detailed world models. It is probable that individuals with BN have a disproportionate reliance on habit-forming processes (Berner & Marsh, 2014). For example, binge-eating and purging episodes occur repeatedly, and are typically prompted by specific stimuli, such as negative affect (Haedt-Matt & Keel, 2011). For instance, an individual with BN might experience guilt or shame after eating something that is deemed forbidden. A binge/purge cycle may ensue as an attempt to cope with that negative emotion. Once learned, this cycle may become perpetuated in a habitual manner. Though no studies have conclusively determined the influence of habitual-control (model-free learning) on BN pathophysiology, some studies show reduced utilization of model-based control in subclinical eating disorder populations (Gillan et al., 2016), as well as clinical populations with anorexia nervosa (Foerde et al., 2019).

1.2.2 Learning Rate

Learning rate is also an important aspect of reward-related decision-making. This refers to the ability to update expectations according to prediction error (the difference between an outcome and a prior expectation regarding that outcome; Addicott et al., 2017). The reward prediction error can be characterized by a temporal difference (TD) algorithm (Sutton & Barto, 1981), which is commonly accepted as an adequate description of the learning process. During TD learning, an agent updates their predictions about the

environment in successive time-steps, and before the final outcome is known. Learning occurs when there is a change in predictions over time. The TD model (Sutton & Barto, 1981) can be mapped onto regions of the brain that are associated with reward-learning, and the characteristics of its signal correspond to dopamine (DA) prediction error-signalling (Montague et al., 1996). Results from Frank et al. (2011) indicate that individuals with BN have significantly weaker brain response in the bilateral amygdala, insula and left orbitofrontal cortex to computer TD model generated reward values. This suggests individuals with BN may have attenuated reward prediction error signalling, and a reduced learning rate. As a result of this impaired capacity to update the expectations of actions as they become less rewarding over time, those with BN may be less likely to reliably predict the consequences of their actions. This might explain, to some degree, the generally poorer performance on reward learning tasks in BN, and their repeated engagement in maladaptive binge/purge behaviours. For example, individuals with BN may continue to regard binge/purge episodes as emotionally relieving behaviours, even when binging/purging no longer provide relief (i.e., as a result of their low learning rate).

1.2.3 Exploration/Exploitation Trade-off

Another key aspect of decisional strategy is the exploration/exploitation trade-off; described as a process of assessing the balance between selection of the option of highest expected value (exploitation), and exploration of the environment for potentially greater rewards (Auer et al., 2002). In environments with non-stationary reward contingencies, some level of exploration is required since a presently optimal decision strategy may later prove suboptimal. This concept is potentially relevant to BN. For example,

binging/purging may initially serve as an effective strategy for regulating negative emotions (Smyth et al., 2007). However, over time, *exploitation* of this strategy may become increasingly harmful to the individual (i.e., by causing physical, psychological and functional consequences of BN). In contrast, had the individual utilized a more *exploration* based learning strategy, they may have developed more adaptive skills for emotion regulation (e.g., mindfulness).

To our knowledge, there are no published studies of exploration/exploitation balance in BN, although binge eating disorder (BED) participants have been previously shown to exhibit excessive exploratory behaviours (Reiter et al., 2016; Morris et al., 2015). Voon (2015) suggests that this may be a result of a reduced avoidance of uncertainty in the context of losses, rather than a specific tendency toward exploration. In a previous study (yet unpublished data), we found that BN participants exhibited reduced exploratory behaviour during completion of a two-step decision-making paradigm. However, we could not clearly differentiate between exploration that was directed by the degree of randomness in environmental rewards or purely random decisions due to value learning impairments.

1.3 BN AND THE NEURAL BASIS OF DECISION MAKING

1.3.1 Frontostriatal Circuitry

Preliminary research suggests that the aforenoted reward learning deficits in BN patients may result from differences in the frontostriatal circuitry of individuals with BN

compared to healthy controls (Berner & Marsh, 2014). Research by Wagner et al. (2010) indicated that a sample of recovered BN women demonstrated altered striatal response to reward on a monetary task, and difficulty in responding to positive and negative feedback. Skunde et al. (2016) showed that frontostriatal *hypoactivity* may specifically contribute to dysregulated reward processing and abnormal eating behaviour in those with BN. On a neurotransmitter level, dysregulated dopaminergic pathways in frontostriatal circuits are thought to be altered in individuals with BN (Michaelides et al., 2012; Berner & Marsh, 2014).

1.3.2 General Dopamine Findings in BN

DA is a neurotransmitter that is essential to food-based motivation and food reward (Baptista, 1999). Research shows that DA metabolites are often reduced in BN patients in comparison to healthy controls and are inversely correlated with binge frequency (Jimerson et al.,1992). Further, polymorphisms in dopaminergic genes resulting in lower DA activity, may predispose individuals to heightened expression of traits that co-occur with BN; namely, self-harm, novelty/stimulus seeking and impulsivity (Thaler et al., 2012).

Several studies have investigated the role of catecholaminergic dysfunction in the development of BN behaviour. Grob et al. (2015) found that remitted BN patients developed mild eating disorder symptoms following oral administration of the drug, alpha-methyl-para-tyrosine (AMPT) which induces catecholamine depletion of both DA and norepinephrine stores (Stine et al., 1997). This suggests that individuals with BN are

vulnerable to eating disorder symptoms in response to decreased catecholamine neurotransmission.

1.3.3 Dopamine-Related Reward Learning Findings in BN

1.3.3.1 General Reward Learning: DA Findings

Altered reward-related decision-making has been identified as an important neuropsychological feature of eating disorders, including BN (Wagner et al., 2010; Wu et al., 2016). Disturbances in reward learning may be related to reduced DA-related reward functioning. Frank et al. (2011) showed that BN patients exhibited diminished brain DA responses when learning the associations of visual stimuli and taste rewards. The concept that reward-learning impairments in BN may be related to DA functioning is similarly demonstrated by Grob et al. (2012) who experimentally achieved catecholamine depletion with AMPT in remitted BN participants. The study found that, under catecholamine depletion, the remitted BN group showed reduced reward learning on a probabilistic reward task compared to healthy controls. No significant differences in reward-processing were observed between the remitted BN participants and healthy controls in the placebo condition. The study findings demonstrate the potential for DA-related learning deficits in BN.

Catecholaminergic dysfunctions in reward learning may be modulated by the neurotrophin brain derived neurotrophic factor (BDNF). Preliminary findings by Homan et al. (2015), showed that AMPT induced differences in plasma BDNF were positively correlated with AMPT related differences in reward-learning for individuals with

remitted BN. The study suggests a relationship between BDNF and DA in reward learning among remitted BN patients.

1.3.3.2 Goal-Directed and Habitual Control Related DA Findings

To date, little is known about the effect of DA on goal-directed and habitual control in those with BN. However, studies have investigated the role of DA among other patient populations. Overall, these studies indicate that DA depletion is associated with decreased goal-directed behaviour. De Wit et al. (2012) found that dopaminergic deficits among Parkinson's disease patients were associated with impaired goal-directed action. Pharmacological enhancement of DA with Levodopa has also been shown to increase goal-directed control in Parkinson's disease patients (Sharp et al., 2016) and in healthy controls (Wunderlich et al., 2012). It is feasible that the effect of DA in the arbitration between goal-directed and habitual action is similar among individuals with BN.

1.3.3.3 Learning Rate Related DA Findings

The effect of DA on learning rate in BN populations is poorly understood. However, research conducted among healthy adult populations shows that DA is relevant to a related concept known as reward prediction error signalling. This refers to a teaching signal that is used to calculate the difference between predicted and actual environmental outcomes (Wang et al., 2020). Simply put, the prediction error signal is a neurological mechanism that helps a person learn when an outcome is different than initially expected. Studies show that DA neurons parallel this signal and report on the difference between expected and observed reward values (Schultz et al., 1997; Cohen et al., 2012). It is

possible that individuals with BN experience unreliable prediction error signalling resulting from presumed deficits in DA functioning. As learning rate scales the reward prediction error signal to govern its influence on learning, it is likely that BN patients experience impaired learning efficiency relative to healthy populations.

1.3.3.4 Exploration/Exploitation Trade-Off Related DA Findings

DA may also be implicated in the exploration/exploitation trade-off. According to several recent computational studies, the functional polymorphism, Val158Met, in the catechol-O-methyltransferase (COMT) gene, is associated with variability in exploratory behaviour (Kayser et al., 2015). As COMT enzyme activity degrades synaptically produced dopamine, those presenting with the less active polymorphism (methionine/ methionine) (Met/Met) are thought to demonstrate greater exploration than individuals with higher activity alleles (Frank et al., 2009). Although no studies have investigated the dopaminergic basis of exploratory/exploitatory behaviour in BN, it is conceivable that reduced DA activity among individuals in this group would limit their exploratory behaviour.

1.4 Reinforcement Learning Task

As discussed, BN symptoms may be governed by reward-based decision-making that is mediated by the arbitration between goal-directed and habitual control, exploratory and exploitatory action, and learning rate. We aimed to explore these mechanisms in individuals with BN using a computational, reinforcement learning framework.

Specifically, a two-step reinforcement learning task was used to quantify the degree to which participant action selection was influenced by these three factors. Furthermore, we were interested in whether these reinforcement learning factors were affected before, during and after treatment with lisdexamfetamine dimesylate (LDX).

1.5 LISDEXAMFETAMINE DIMESYLATE

Lisdexamfetamine dimesylate (LDX; Vyvanse) is a prescription medication used in the treatment of attention-deficit/hyperactivity disorder (ADHD) in children, adolescents, and adults. There is also significant evidence to support the use of LDX for moderate to severe BED in adults. In a phase II randomized, placebo-controlled clinical trial investigating the efficacy of LDX for BED, it was found that those treated with 50 mg/day or 70 mg/day dosages experienced a significant reduction in binge episodes and binge days from baseline to week 11 of treatment. Compared with the placebo group, those treated with LDX also experienced greater binge eating cessation and global improvement in BED symptoms (McElroy et al., 2015). Two identically designed phase III trials produced similar study findings, leading to FDA approval of LDX for moderate to severe BED treatment in 2015, and ultimately, Health Canada approval in 2016.

Preliminary case reports suggest that LDX may also be effective for treating BN patients who had not previously responded to other forms of pharmacotherapy and psychotherapy (Keshen & Helson, 2017). In a 2021 feasibility study examining the potential efficacy of LDX in adults with BN, Keshen et al. also found that LDX use resulted in a clinically

significant reduction in binge episodes and compensatory behaviours from baseline to end-of-treatment. Little is known about the neural mechanisms by which LDX can improve the symptoms of BN. However, previous studies suggest that LDX can improve BED symptoms by normalizing frontostriatal activation via increased DA transmission in this region (i.e., the same reinforcement learning related brain regions thought to be relevant in the pathoetiology of BN; Griffiths et al., 2019).

1.6 Study Aims/ Hypotheses

Study data was used to address the questions: A) Does dopaminergic medication affect exploration/exploitation balance, learning rate and goal-directed control among individuals with BN? B) Can LDX use produce a clinically meaningful response (i.e., reduction in number of binge/purge episodes) that is related to these aspects of reinforcement learning?

We hypothesized that participants would demonstrate increased exploration, learning rate and goal-directed control after the initiation of LDX. Further, we hypothesized that this would correlate with a decrease in binge/purge behaviours. All changes were expected to occur in a dose-related fashion.

CHAPTER 2 METHODS

The study occurred as a sub-project of an existing feasibility study by Keshen et al. (2021) examining the effectiveness of LDX in BN patients.

2.1 PARTICIPANTS

Participants were recruited from the Greater Halifax Area in Nova Scotia, Canada through online advertisements on Facebook, Instagram and Kijiji. Paper recruitment handouts were also distributed at relevant locations including university health centres and psychology clinics. To determine eligibility for a screening visit, interested participants completed an online pre-screening questionnaire (Appendix A). Potential participants were called by an investigator for a preliminary screening interview. Those eligible for study participation were required to attend an in-person screening session at the Nova Scotia Health Eating Disorder Program (Appendix B). The criteria used in the evaluation of participant inclusion/exclusion is presented in Appendix C.

In the original feasibility study by Keshen et al. (2021) it was determined that a sample of 30 participants would allow for a reasonable probability of having a minimum of 20 completers at follow-up. This sample size was based on a dropout rate of 15-30% that was observed in comparable studies (McElroy et al., 2015) and was deemed appropriate to measure the feasibility parameters in the primary study. Power calculations were not completed for this sub-project because the analyses were exploratory and hypothesis

generating. Enrollment commenced in September 2018 and was temporarily paused in March 2020 due to Coronavirus disease (COVID-19) related research restrictions. Study enrollment was resumed in June 2020 but was permanently closed in July 2020, because of COVID-19. This resulted in a total sample of 23 participants, instead of the original aim of 30. Given that our study occurred as a pilot project, this sample size was adequate to measure possible associations that would be worth exploring in future studies.

Of the 23 participants enrolled in the study, 18 completed the study per protocol. One participant was withdrawn from the study for noncompliance, and one participant was withdrawn for a loss of greater than 5% body weight within a given month (see Appendix D for the criteria for study discontinuation). Three participants dropped out of the study prematurely.

2.1.1 Participant Compensation

Participants were compensated $20.00 following each study session (total of 5 sessions) for study-related parking and travel costs. Those withdrawing from the study early were not eligible to receive further compensation**.**

2.2 Intervention: Study Drug

Study participants were administered LDX. The trial began with a 4-week titration period followed by a 4-week maintenance period. The drug was initially administered to each participant at 30mg/day and was increased weekly by 20mg increments until the optimal dose was achieved (50mg/day or 70mg/day), as determined by the principal investigator.

The final week of the titration period was designated for a drug dosage reduction (from 70mg/day to 50mg/day), if the participant was found to be intolerant of the maximum dosage. LDX administration ended at week 9 of the study. A weekly medication administration schedule is presented in Appendix E with accompanying study procedures.

2.3 Measures And Materials

2.3.1 Contextual Bandit Reinforcement Learning Task

Participants completed 250 trials of the task (plus any additional resulting from aborted trials), in two blocks separated by a 30 second break. The first block consisted of 84 trials, with the remainder to be completed in the second block. In the first stage, participants selected between one of two pairs of spaceships. These pairs were essentially equivalent at their first state. Each selection led deterministically, to one of two second-step planets. On each planet, participants were presented with an "alien" that "mined" from a "space mine". Mine payoffs resulted in either the presentation of reward in the form of "space treasure", or the omission of reward in the form of "antimatter" and fluctuated according to Gaussian random walk. This refers to a mathematical object that consists of successive steps along a normal distribution. At the end of the experiment, participants were given 1¢ for every two points they earned on the task. The original protocol for the contextual bandit reinforcement learning task was developed by Kool et al. (2016). The implementation of the task, written in JavaScript, was directly translated

into the PsychoPy framework by Dr. Abraham Nunes and Alexander Rudiuk. A detailed description of the task is presented in Appendix F.

Prior to the main task, participants received extensive task training. They were presented with a "storyline" of the task and were familiarized with task icons/images. Participants were instructed on the process of obtaining rewards and learned about the state-transition structure of the task. All participants were required to undertake 25 full practice trials before performing the task. If a trial was aborted due to computer time-out, an additional trial was added at the end of the practice session.

The task was implemented in our study for the purpose of assessing an individual's relative degree of decisional, goal-directed versus habitual behavioural control, learning rate and exploration/exploitation. Goal-directed (model-based control) was assessed by evaluating participants' reliance on a deliberative method of action selection that acknowledged the causal structure of the task. Those who relied on habitual (model-free) decision-making would not consider an explicit causal model of the task and would be more likely to simply repeat actions that were previously rewarded.

Participant learning rate was measured by evaluating the degree to which participant choices were updated by prediction error. Those with a higher learning rate would be more sensitive to the most recent value of each action on the task and would be less reliant on prior beliefs during action-selection.

Last, we intended to measure exploration and exploitation using a parameter called inverse softmax temperature. A higher inverse softmax temperature would reflect a tendency toward the selection of options with highest previous payoff (exploitation). Dissimilarly, a lower inverse softmax temperature would reflect a tendency to deviate from this behaviour (exploration) (Addicott et al., 2017).

2.3.2 Task Quality Assurance

In October 2016, Dr. A. Nunes conducted pilot testing of the task. A group of 12 individuals who internally tested the task indicated that the task was clear and understandable. One tester noted that the original task graphics by Kool et al. (2016), were not distinguishable for colour-blind individuals. Given this feedback, the task was modified to incorporate a colour-blind friendly palette. This occurred prior to implementing the task in our study.

Pilot testing was also used to verify the reliability of the software across machines. Task developers confirmed that the task rendered in full resolution on Windows 7, Windows 10, MacOS and Ubuntu Linux. The task was found to operate reliably, and without disruption on the study computer located at the Nova Scotia Health Eating Disorder Program. We also ensured that the task functioned appropriately in uncovering the underlying psychological functions of interest by testing how accurately the reinforcement learning parameters of the model could be estimated.

2.3.3 Clinical Response Measures

Eating Disorder Examination 17.0D (EDE 17.0D): Participants were administered the EDE 17.0D by trained interviewers. This was used to gather information on participant objective binge episodes/days and subjective binge episodes/days over the previous 28 days. The EDE 17.0D was also used to determine the number of episodes/days that participants engaged in self-induced vomiting, laxative misuse, and diuretic misuse over the 28-day period. Test-retest reliabilities of the EDE 17.0D are presented in Appendix G.

Dietary Records: Participants were asked to keep a record of all food and beverages consumed. They were required to self-report episodes of binge-eating, self-induced vomiting, laxative-taking and diuretic-taking.

2.3.4 Covariate Measures

Operation Span Task (OSPAN) (computer-based): The operation span task (OSPAN) is a paradigm for the assessment of working memory (WM) (Conway et al., 2005). The existing implementation by Titus von der Malsburg (https://github.com/tmalsburg/py-span-task) was translated into the context of PsychoPy behavioural task development package by Dr. A. Nunes in September-October 2016 for use in this study. Test-retest reliabilities of the OSPAN task are presented in Appendix G.

During the OSPAN task, participants were required to validate simple arithmetic equations and memorize consonants that appeared after each item. At the end of each set of equations, participants were prompted to recall the list of consonants in the correct

serial order. The average length of the lists that could be recalled was the participant's operation span.

OSPAN administration was relevant to understanding participants' model-based and model-free contributions to task learning under stress. Previous research suggests that neuropsychological stress response diminishes the contribution of model-based choice in individuals with low WM capacity, but not high WM capacity (Otto, Raio et al., 2013). OSPAN administration would allow us to determine the potential influence of acute stress during task performance on participant WM, and to examine if differences in WM capacity influenced participants' reliance on model-based and model-free responding.

As striatal DA is also critically implicated in WM functioning (Bäckman et al., 2017), it is likely that participants would experience WM improvement following LDX administration. This would encourage model-based planning during task administration.

2.4 Procedures

The single-site trial was conducted out of the Nova Scotia Health Eating Disorder Program (Abbie J Lane Building; QEII Health Sciences Centre) in Halifax, Nova Scotia, Canada.

The contextual bandit reinforcement learning task and OSPAN task were administered to participants on a computer located at the study site at week 1 (baseline), week 2, week 9

(maintenance dose), and week 10 (follow-up). The total duration of these tasks was approximately 45 minutes. Based on the pharmacokinetics of LDX, we administered computer tasks 4-8 hours after participants ingested the study medication.

Dietary records were collected from participants weekly throughout the course of the study. The EDE 17.0D was administered to the participants at the Nova Scotia Health Eating Disorder Program at baseline (week 1) and week 9. Where possible, participant responses to EDE questions regarding the previous 28 days were verified with information from their dietary records. A detailed patient schedule is presented in Appendix E.

2.5 Statistical Analysis

Analysis of participant behavioural data was performed using *theory-free* and *theory-based* methods. Our *theory-free* approach involved the use of permutation testing for paired comparisons. This approach was used to determine if the distributions of participant BN symptom data were different, before and during LDX treatment. The approach was also used to compare differences in participant reward rate on the two-step task at various dosages of LDX. Reward rate was calculated by dividing the number of points received on the task by the total number of task trials (250 trials). No assumptions were made about the strategy that participants used while performing the task.

Our *theory-based* portion of analysis was not used to detect an effect, but rather explain an effect. Specifically, computational learning models were fit to subjects' trial-by-trial behaviours to quantify (A) the likely decision-making strategy being used and (B) the degree to which specific model parameters, such as model-based/model-free control, exploration-exploitation balance, and learning rate were used at each time point. This was performed by constructing reinforcement learning models that mirrored the participants' trial-by-trial behavioral data.

2.5.1 Paired Permutation Testing

We used R (R Core Team, 2020), and the wPerm package (Weiss, 2015) to perform permutation testing for non-independent matched pair data. This was used to compare participant behavioural outcomes at different time points throughout the study. Nonparametric permutation testing was used over alternative parametric approaches because our experimental data did not satisfy the statistical assumptions underlying traditional, parametric tests. For example, the sample data were not all normal in form, and were not all of equal variance. As permutation tests make no distributional assumptions, they were useful in making inferences about the location of study data at different timepoints. We used an alpha level of .05 for all permutation tests.

2.5.2 Computational Modeling of Reinforcement Learning

Computational modeling analysis proceeded in two phases: model selection, and parameter estimation. Analyses were performed by Dr. Abraham Nunes in the Python programming language. Model parameters were estimated using expectation-

maximization (Huys et al. 2011), and model comparison done using Bayesian model selection (Rigoux et al. 2014) at each time point. We present only the parameters of the model with the highest probability given by the Bayesian model selection procedure.

In the model selection phase, reinforcement learning models (Appendix H) were fit to each participant's trial-by-trial behavioural data. Fitting the models to the participants' behavioural data yielded an $n_{subject} \times n_{model}$ matrix of approximations to the logarithmic model evidence, which was then submitted to the Bayesian Model Selection procedure (Rigoux et al., 2014) in order to identify the most probable model that explained the aggregate group's behavioural data.

The models investigated were a model-free reinforcement learning algorithm developed using State-Action-Reward-State-Action, SARSA (λ) temporal difference learning (Sutton, 1998, Sutton et al., 1999), a model-based reinforcement learning algorithm using Bellman equation (Bellman, 1957), and a hybrid model with model-based and model-free subcomponents. In the hybrid model, the relative weighting of participant model-based and model-free strategies was parameterized by *w,* where 1 indicates pure model-based learning, and 0 denotes pure model-free learning (Appendix H).

The definitions of these functions are as follows:

- α: Learning rate: a coefficient which indicates how quickly an agent updates their state-action reward expectations. Participants with a high learning rate would be able to acquire reward related information about their actions quickly.

- β: Inverse softmax temperature parameter (an index of the amount of choice randomness). When $\beta \rightarrow 0$, choice consistency decreases and actions become more random.
- ρ: Perseveration: The tendency to repeat a previously selected action regardless of the action's value. Participants with high perseverance, would tend to repeat choices made on the previous trial of the task.

2.5.3 Disentangling Medication Effects from Practice Effects

Changes in performance due to repeated exposure to test items (practice effects) were addressed. Participants were administered computer-based tasks at week 1 (baseline) prior to beginning LDX treatment. Task results were recorded at week 10 (follow-up), to determine if participant performance returned to baseline levels after LDX discontinuation (non-indicative of practice effects), or if performance was greater at follow-up relative to baseline (indicative of practice effects). A similar method of dissociating medication effects from practice effects was demonstrated by Boulay et al. (2007) who assessed for practice effects on neurocognitive functioning, before and after randomization to treatment.

CHAPTER 3 RESULTS

3.1 DEMOGRAPHICS

Demographics of the sample are summarized in Table 1. 23 female participants with moderate-extreme BN were enrolled in the study (*M*=26.83 years, *SD*=7.96 years). Of these, 22 participants (95.65%) identified as Caucasian, and 1 participant (4.35%) identified as Caucasian / First Nations. 15 participants (65.22%) had engaged in a previous treatment for BN and 9 participants (39.13%) were taking a psychiatric medication other than LDX throughout the duration of the study.

Table 1

Demographics of Study Sample

Baseline Characteristic	*n*	*%*	*M*	*SD*
Age			26.83	7.96
Baseline BMI			24.53	2.54
Gender				
Female	23	100.0		
Race:				
Caucasian	22	95.65		
Other: Caucasian / First Nations	1	4.35		
Education:				

Baseline Characteristic	*n*	*%*	*M*	*SD*
High school diploma	1	4.35		
Some college credit, no degree	2	8.70		
Some university, no degree	6	26.09		
Trade/technical/vocational training	2	8.70		
Bachelor's degree	7	30.43		
Master's degree	1	4.35		
Other	4	17.39		
Marital Status:				
Single	18	78.26		
Married	3	13.04		
Other	2	8.70		
Employment Status:				
Employed, full-time	11	47.83		
Employed, part-time or casual	2	8.70		
Employed, part-time or casual & student	3	13.04		
Student	5	21.74		
Unemployed	2	8.70		
Illness Duration (years)			9.33	7.73
Previous Tx attempt for ED				

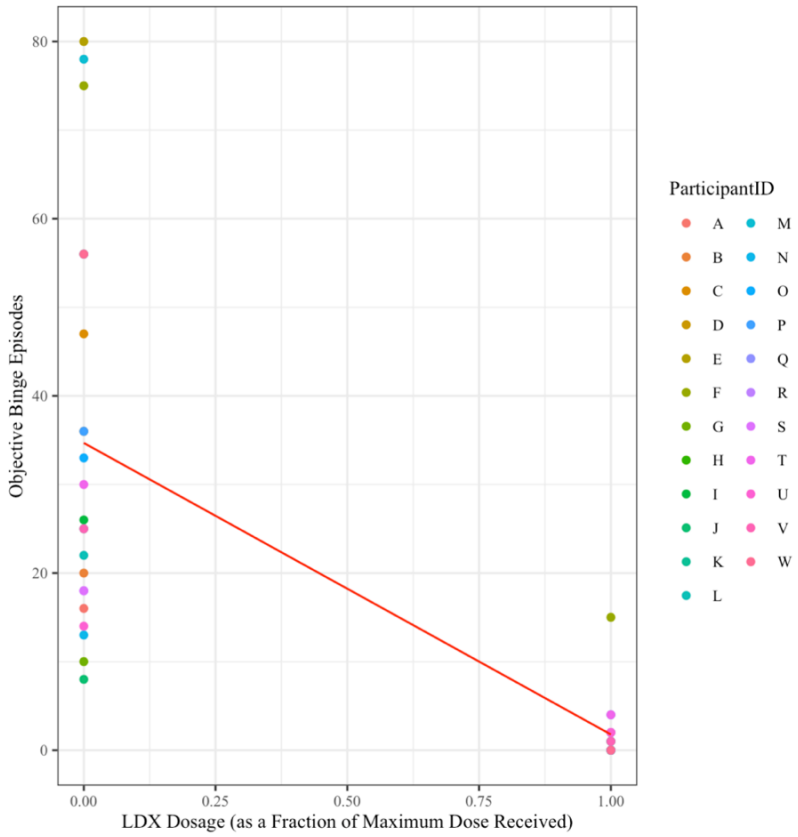
Baseline Characteristic	*n*	*%*	*M*	*SD*
No	8	34.78		
Yes	15	65.22		
Current Psych Medication				
No	14	60.87		
Yes	9	39.13		

3.2 Paired Samples Permutation Tests

Results of permutation paired location tests based on 9999 replications are presented in Appendix I. These results indicate that the distribution of objective binge episodes at T3 (maintenance dose) had significantly smaller values than objective binge episodes at T1 (baseline) (M_{T3}-M_{T1}= -35.11) ($p < 0.001$). This is similarly shown in Figure 3.1, which shows that participant objective binge episodes were reduced after LDX administration.

Figure 3.1

Scatterplot of Participant Objective Binge Episodes Versus LDX Dosage



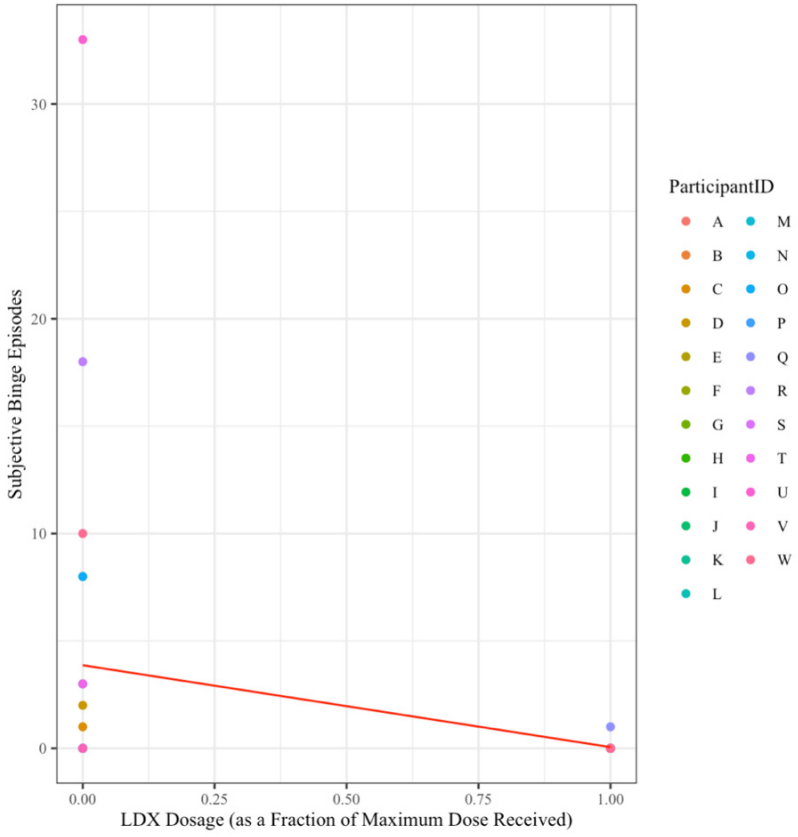
Note: Line connects the means of grouped data at various dosages of LDX (as a fraction of the maximum dose received).

The frequency of objective binge days was also found to be lowest when LDX was administered at maintenance dosage. Paired permutation tests show that the frequency of objective binge days at T3 was lower relative to objective binge days at T1 (M_{T3}-M_{T1}= -16.84) ($p < 0.001$).

We conducted additional testing on the effect of LDX on participant subjective binging. To establish whether LDX use resulted in a reduction of subjective binge episodes we performed paired permutation testing of the distributions of participant subjective binge episodes at T1 and T3. The tests indicate that subjective binging was significantly lower at T3 than at T1 (prior to beginning LDX) (M_{T3}-M_{T1}= -3.579) (p= 0.0064). Figure 3.2 provides a visual representation of this association.

Figure 3.2

Scatterplot of Participant Subjective Binge Episodes Versus LDX Dosage



Note: Line connects the means of grouped data at various dosages of LDX (as a fraction

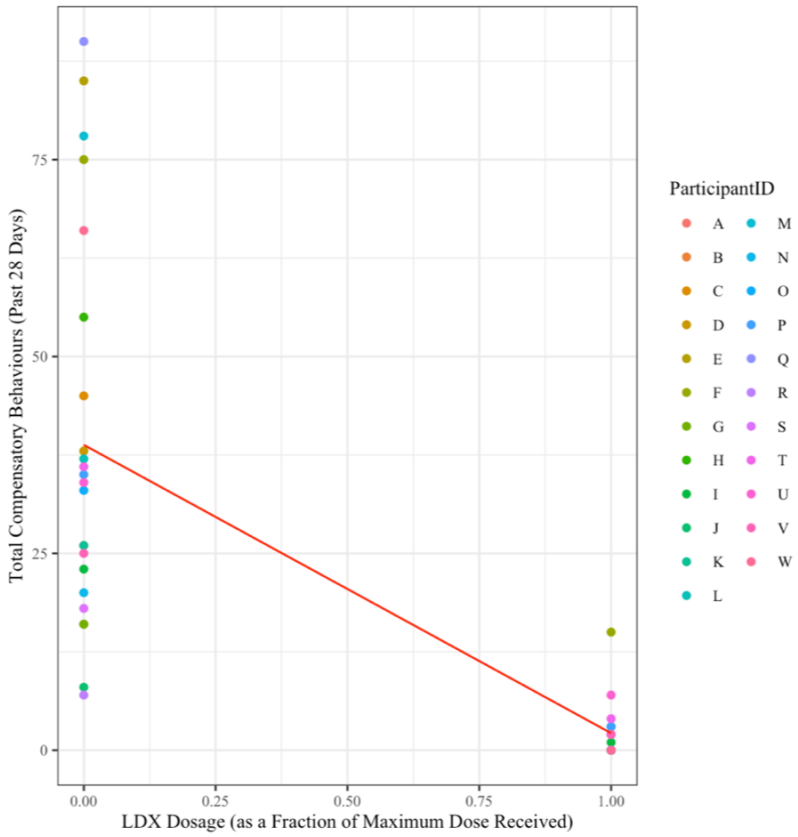
of the maximum dose received).

Similar to the results from the permutation tests describing the distribution of participant subjective binge episodes at T1 and T3, participant subjective binge days were also found to be significantly lower at T3 (M_{T3}-M_{T1}= -3.105, p= 0.0059).

It was also necessary to examine the distribution of values for participant compensatory behaviours before, and during LDX treatment. It was determined that compensatory behaviours at T3 were systematically less than at T1 (M_{T3}-M_{T1}= -38.85, p< 0.001). Figure 3.3 indicates that participants engaged in compensatory behaviours most frequently, prior to LDX administration. Compensatory actions were reduced with increasing LDX use.

Figure 3.3

Scatterplot of Participant Compensatory Behaviours Versus LDX Dosage



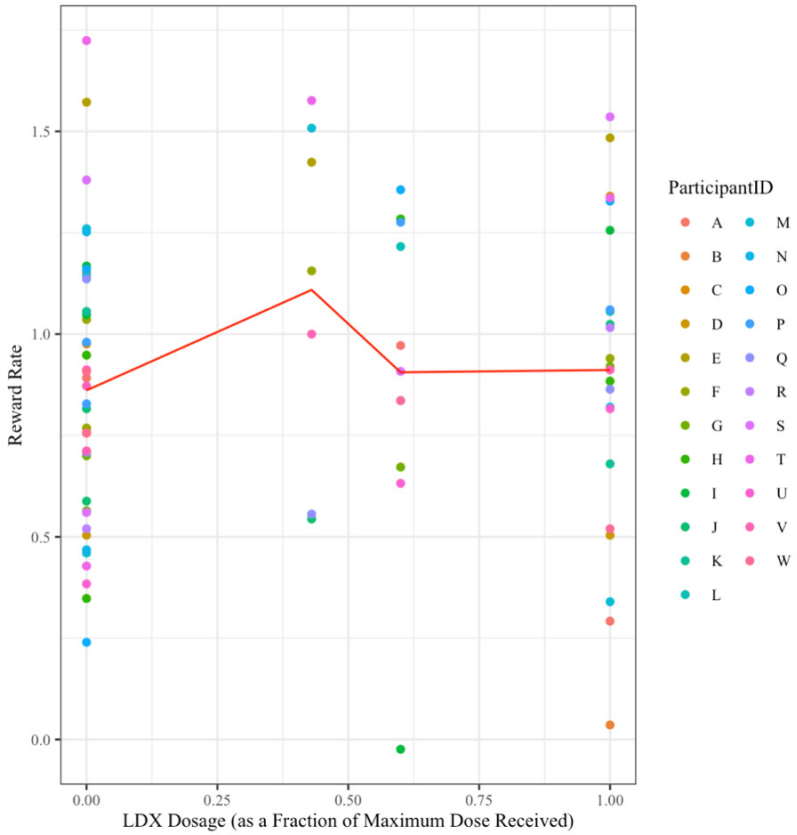
Note: Line connects the means of grouped data at various dosages of LDX (as a fraction of the maximum dose received).

We aimed to determine whether LDX had the ability to alter participant reinforcement learning, and if changes in participant learning were related to the observed reduction in BN symptoms. Raw data of participant reward rate on the two-step task versus LDX dosage (Figure 3.4) shows that participants did not experience a significant increase in reward rate with increasing medication use. This is supported by the results of paired

permutation testing which indicate that participant reward rate was not significantly greater at T3 than at T1 (M_{T3}-M_{T1}= 0.1511, p=0.0702). Similarly, participant reward rate at T3 was not systematically greater than participant reward rate at T4 (off study medication) (M_{T3}-M_{T4}= 0.0586, p=0.313).

Figure 3.4

Scatterplot of Participant Reward Rate Versus LDX Dosage



Note: Line connects the means of grouped data at various dosages of LDX (as a fraction of the maximum dose received).

Last, the distributions corresponding to WM were compared among time points. Two-tailed paired permutation tests indicate that the distribution of OSPAN scores while on the maintenance dose of LDX (T3) did not have systematically smaller or larger values than OSPAN scores at T1 (M_{T3}-M_{T1}= -0.9286 p=0.887).

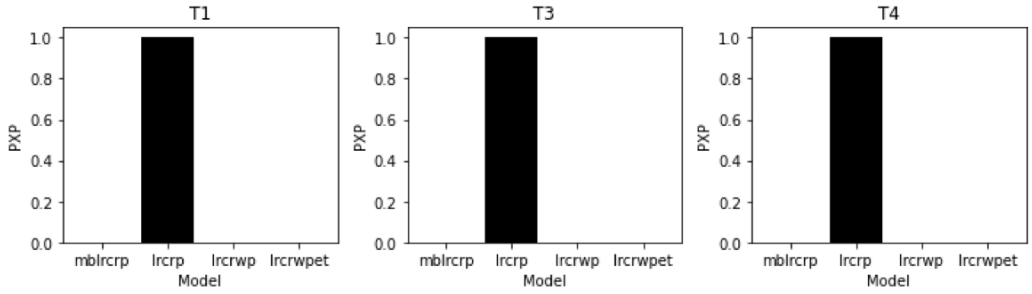
3.3 Theory-Based Analysis

3.3.1 Model Selection

At time points 1, 3, and 4, reinforcement learning models were fit to each subject's trial-by-trial behavioural data. These results suggest that at all time steps, the aggregate group's data were best explained by a model-free reinforcement learning approach (Figure 3.5).

Figure 3.5

Bayesian Model Selection Results



Note: mblrcrp= Model-based agent with learning rate, inverse softmax temperature, and perseveration; lrcrp= Model-free agent with learning rate, inverse softmax temperature,

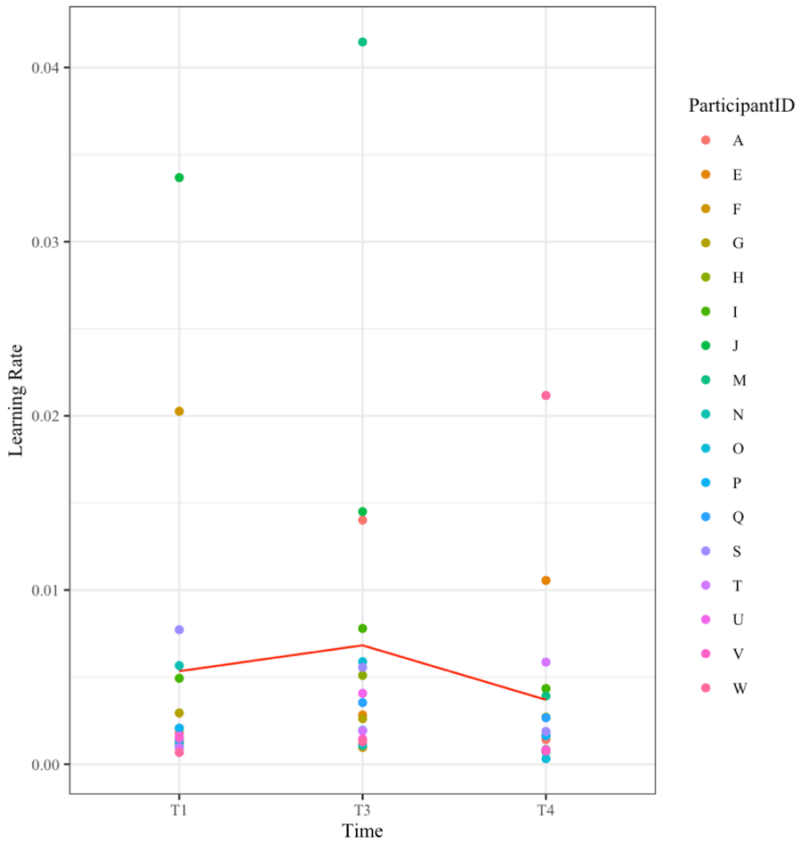
and perseveration; lrcrwp=Hybrid agent with learning rate, inverse softmax temperature, and perseveration; lrcrwpet=Hybrid agent with learning rate, inverse softmax temperature, perseveration, and eligibility trace; pxp=protected exceedance probability.

3.3.2 Parameter Estimation

A table of estimated model parameters at T1, T3, and T4 are presented in Appendix J. Participant learning rate was close to 0 across all time points, indicating no learning was taking place. This is illustrated in Figure 3.6, in which participant learning rate estimates using the model-free model are plotted at different timepoints. Descriptive statistics of participant learning rate are presented in Table 2.

Figure 3.6

Estimation of Learning Rate at Various Time Points (Model-Free Agent).



Note: Line connects the means of grouped data at various time points.

Table 2

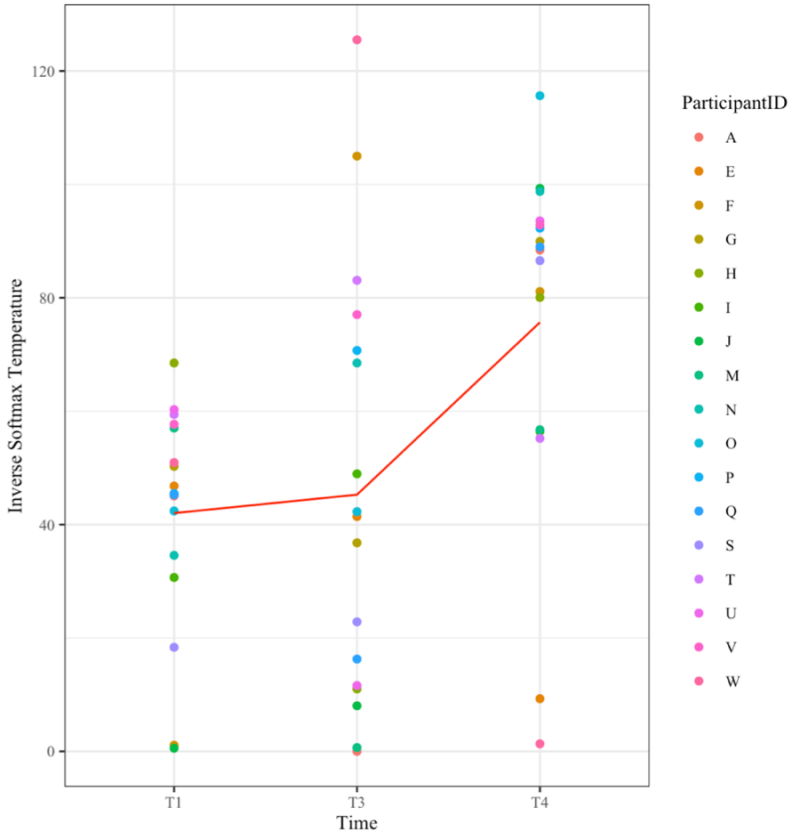
Descriptive Statistics: Learning Rate

Time	Count	*M*	*SE*	*SD*	Min	Max
1	17	0.0053	0.0021	0.0087	0.0007	0.0337
3	17	0.0068	0.0024	0.0098	0.0010	0.0415
4	17	0.0037	0.0013	0.0052	0.0003	0.0212

Participant inverse softmax temperatures were also mostly high (Appendix J). Typically, a higher inverse softmax temperature could reflect a more exploitatory choice strategy, however, as the participants were insensitive to reward contingencies, discussion of inverse softmax temperature is irrelevant in this instance. Figure 3.7 provides a scatterplot of estimated inverse softmax temperature values at T1, T3, and T4. Descriptive statistics are included in Table 3.

Figure 3.7

Estimation of Inverse Softmax Temperature at Various Time Points (Model-Free Agent).



Note: Line connects the means of grouped data at various time points.

Table 3

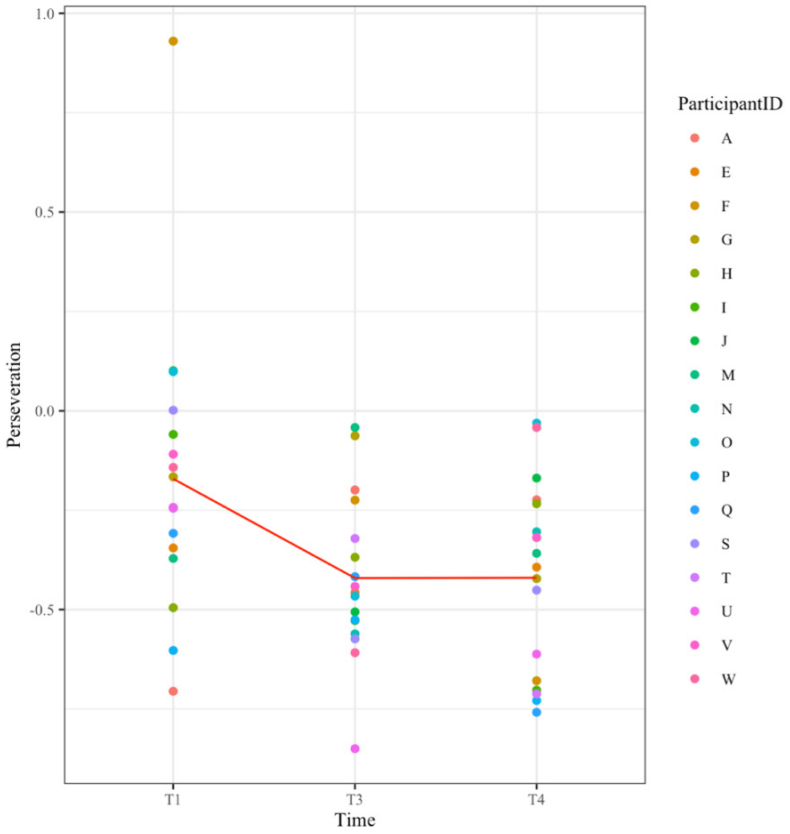
Descriptive Statistics: Inverse Softmax Temperature

Time	Count	*M*	*SE*	*SD*	Min	Max
1	17	42.03	4.755	19.60	0.5628	68.50
3	17	45.28	9.158	37.76	0.0104	125.5
4	17	75.67	7.517	30.99	1.336	115.6

Finally, perseveration values were low and negative across all time points (Table 4). Ordinarily, this would indicate that participants shifted readily between choices on the task. However, as participants were unable to learn the task reward values, it is unimportant to discuss their reliance on perseverative action. Estimated perseveration values at T1, T3 and T4 are plotted in Figure 3.8.

Figure 3.8

Estimation of Perseveration at Various Time-Points (Model-Free Agent).



Note: Line connects the means of grouped data at various time points.

Table 4

Descriptive Statistics: Perseveration

Time	Count	*M*	*SE*	*SD*	Min	Max
1	17	-0.17094	0.0876	0.3612	-0.7055	0.9299
3	17	-0.4208	0.0496	0.2043	-0.8500	-0.0420
4	17	-0.4200	0.0588	0.2422	-0.7585	-0.0308

CHAPTER 4 DISCUSSION

Research indicates that deficits in reward learning are implicated in development of binge/purge behaviours (Wagner et al., 2010). Though BN pathophysiology is poorly understood (Kekic et al., 2016), these disturbances in reward processing are thought to be related to DA neurocircuits (Grob et al., 2012). The purpose of this study was to assess the effect of LDX administration (and presumed increases in DA transmission) on participant goal-directed and habitual responding, exploration/exploitation balance and learning rate. These distinct subcomponents of reinforcement learning were measured using a two-step reinforcement learning task in which participants made a series of choices between two stimuli. Each choice deterministically transitioned to a second-stage state that was associated with a fluctuating reward payoff. We wanted to determine the effect of LDX on the reinforcement learning functions of interest, and to assess whether changes in these elements were related to a clinically meaningful response (i.e., decreases in binge/purge frequency).

Results from paired permutation tests are presented first and compare differences in BN symptom data and reward rate before and during LDX treatment. A discussion of the computational learning model that best explains the aggregate group's behavioural data follows. Subsequently, an explanation of model parameters including learning rate, inverse softmax temperature, and perseveration is provided. Later, study strengths and limitations are discussed. This is followed by a description of study implications and recommendations for future research.

4.1 Study Findings

4.1.1 Paired Permutation Testing

Results indicate that study participants experienced a reduction in BN symptoms during LDX treatment. Permutation tests for repeat measures data showed that the distributions of participant objective binge episodes (M_{T3}-M_{T1}= -35.11, $p < 0.001$) and objective binge days (M_{T3}-M_{T1}= -16.84, $p < 0.001$) were less at T3 (maintenance dose) than at T1 (baseline). The distribution of total compensatory behaviours was also found to be significantly less at T3 relative to T1 (M_{T3}-M_{T1}= -38.85, $p< 0.001$). This large and statistically significant reduction in participant objective binging and compensatory behaviours is likely to have resulted in substantial improvement in the participants' quality of life. This decrease in BN pathology was expected with LDX administration and is consistent with the results of a case-report study in which LDX use resulted in a reduction in binge/purge days per month in BN participants, one month following the medication initiation (Keshen & Helson, 2017). In a previous study by McElory et al. (2015), LDX administration also produced a statistically significant decrease in binge eating days per week relative to placebo in a group of individuals with moderate to severe BED.

While a reduction in BN symptoms occurred with LDX treatment, the goal of this sub-project was to determine if this was related to changes in participant reinforcement learning. Results from paired permutation testing show that the distribution of participant reward rate at T3 was not significantly greater than at T1. This indicates that participants were not able to learn the reinforcement task at all, regardless of increasing LDX use.

This was an unexpected result, as we had originally hypothesized that an increase in DA activity as a result of LDX use would increase participant reinforcement learning. It is possible that participants' low reward rate was related to their use of an inflexible model-free learning strategy. For example, participants might have been rewarded less frequently because they habitually selected previously rewarded actions instead of acknowledging the causal model of the task. This finding is supported by previous research by Kool et al. (2016), who found that measures of model-based planning during the reinforcement learning task, showed a positive correlation with reward rate.

Last, two-tailed permutation tests were used to compare participant WM capacity before, and after LDX administration. Results indicate that the distribution of participant OSPAN scores at maintenance drug dose was neither systematically smaller or larger than OSPAN scores at baseline. We had originally expected that LDX administration would result in a marked improvement in participants' WM performance, as prior research suggests that psychostimulant treatment may have beneficial effects on WM capacity (Wong & Stevens, 2012). However, whether differences in LDX use can result in an improvement in WM among adults with BN remains unknown.

It is also possible that deficits in WM capacity were unrelated to the participants' level of BN symptoms. For example, it is feasible that participants experienced a reduction in binge/purge behaviour without a corresponding increase in WM. Research by Barnett et al. (2001) indicated that a group of children receiving psychostimulant treatment for ADHD had improved WM relative to non-medicated children with ADHD, but that the

magnitude of their ADHD symptoms was unrelated to their WM. A study by Salmi et al. (2020) also indicated that psychiatric symptoms and self-rated cognition in a group of Parkinson's Disease patients were weakly linked to WM performance. As BN, ADHD and Parkinson's Disease share common neurobiological features including dysregulated DA signalling, it is possible that WM has a similar, limited role in the clinical manifestation of BN.

4.1.2 Model Selection

Model comparison was performed to determine which of the possible models best fit the aggregate group's behavioural data. At all time points participants used model-free (habitual) control during the completion of the two-step task. Under this purely model-free strategy, participants strengthened or weakened associations between stimuli and actions, depending on whether the action was followed by a reward or not (Sutton & Barto, 1998). Participants were more likely to select previously rewarded actions and would switch actions if they experienced a loss of reward. This simple win-stay lose-switch model free learning strategy (da Silva & Hare, 2019) is in contrast to a purely model-based approach in which participants would compute action values using a model of the task environment.

Originally, we hypothesized that an increase in goal-directed (model-based) control with increasing doses of LDX would be associated with an improvement in BN symptoms. In theory, individuals with greater model-based learning would have more of an awareness of the adverse long-term effects of binging/purging (i.e., dental erosions, periodontal

disease, electrolyte abnormalities and gastrointestinal complications resulting from induced emesis or laxative misuse; Mehler, 2011). Therefore, they would be more likely to reduce their risk of engagement in these behaviours by selecting alternative actions that would help them to manage their negative emotions in a healthier manner. For example, trying a new mindfulness skill or performing breathing exercises to reduce their urge to binge/purge. Our finding that BN participants relied on model-free control only, is unsupported by previous research which shows that an inverse relationship may exist between BN symptom severity and model-based control (Nunes et al., 2018).Voon et al. (2015) also showed that higher binge eating scores in a sample of BED participants were positively correlated with a shift toward model-free (habitual) behaviours. These studies indicate that greater model-based-behaviour, and not model-free behavior may be associated with a reduction in eating disorder symptoms.

While it is unlikely that the reinforcement task lacked ecological validity, this should be acknowledged as a possible reason why performance on the task did not relate to a reduction in BN symptoms. It is feasible that the task does not meaningfully capture the types of decisions that individuals with BN make when they choose to binge or purge. Therefore, it is possible that decision-making on the task cannot be generalized to *real-world* symptom or behavioural changes among BN populations.

Second, it is possible that the lack of improvement in participant model-based control might be related to their lack of improvement in WM. Research indicates that central executive functioning governs deliberative model-based decision-making. Specifically,

the depletion of WM resources can inhibit model-based learning behaviour, and engenders reliance on habitual behaviours (Otto, Gershman et al., 2013). However, it is more probable that participants did not experience an increase in model-based control because they were unable to learn the task's reward contingencies.

Another potential reason that participants were unable to optimize their performance on the task with increasing LDX use is that the medication might not have increased DA receptor density. Though few studies have investigated the role of striatal DA in BN, neuroimaging literature suggests that the "chronically-addicted state" is associated with low levels of striatal DA and reduced DA type 2 (D2) receptors (Volkow et al., 2009). Broft et al. (2011) suggest that preclinical models of "BN-like" eating behaviours also show that decreased D2 receptor density may be implicated in the initiation and perpetuation of BN. As D2 receptors have been shown to contribute to both approach and avoidance learning in healthy adult populations (Jocham et al., 2014) it is possible that the BN participants had impaired reinforcement learning resulting from reduced D2-class receptor density. While LDX can increase synaptic DA concentrations, the drug does not increase D2 receptor site numbers. It is possible that this limited participants' ability to accurately select rewarding actions and reject punishing actions on the two-step task.

It remains unclear how participants experienced a reduction in BN symptoms during treatment with LDX without an increase in reinforcement learning. It is possible that the study medication resulted in participant appetite-suppression that was directly related to their reduction in binge episodes, and relatedly, their decrease in compensatory

behaviours. Research investigating the use of LDX among patient populations with ADHD (Wigal et al., 2011) and BED (Brown et al., 2010) note appetite suppression as a common treatment-emergent adverse event (TEAE) of LDX treatment. This TEAE of LDX treatment is consistent with the effects of long-term stimulant use.

Last, it is feasible that participant BN symptoms were reduced as a result of LDX - induced decreases in emotional lability (EL). EL refers to frequent and intense emotional shifting and is posited to have a role in the maintenance of BN behaviour. Those with BN may engage in dysregulated binge/purge behaviours as a strategy for emotion regulation during periods of uncontrollable shifts in emotional intensity/valence (Anestis et al., 2012). A 2009 study by Anestis et al. found that EL significantly predicted Impulsive Behaviour Scale score in a clinical sample of females with BN, even when controlling for general impulsivity. The results indicate that higher EL among individuals with BN, may increase proneness to destructive behaviour including binging/purging. Prior research suggests that LDX may be effective in reducing EL in specific psychiatric populations. In a double-blind, placebo-controlled group trial investigating changes in EL with LDX administration in a group of children with ADHD, it was found that LDX showed improvement versus the placebo on Conners' Parent Rating Scale items of anger, loss of temper, and irritability (Childress et al., 2014). Given these findings, it is possible that LDX administration reduced the severity of emotional reactivity in the BN patients involved in our study. This decrease in EL would reduce the need for binge eating/purging as emotion regulatory behaviors (Yu & Selby, 2013).

4.1.3 Parameter Estimation

Parameter estimation involved finding the parameter values that best described the participants' task data under a model-free learning strategy. The parameters measured were learning rate, inverse softmax temperature, and perseveration, and provide a succinct summary of the participants' behavioural data on the two-step task. They are also useful for quantifying the effect of LDX administration on participant task-related behaviour, and for evaluating individual differences in participant decision-making (Frank et al., 2007).

Learning Rate

Participant learning rate was negligible across all time points. Moreover, LDX use did not affect the participants' ability to update their understanding of the task (i.e., LDX did not affect learning rate). To reiterate, this was an unanticipated study finding. Research indicates that dopaminergic medications may induce changes in reward prediction error signaling (Diederen et al., 2017). Rutledge et al. (2009) further suggests that this effect may result in differences in the learning rate estimated by standard reinforcement learning models. In a sample of Parkinson's disease patients treated with levodopa, Rutledge et al. (2009) found that learning rate was higher in patients on than off the study medication. It would be expected that the BN participants in our study would experience a similar increase in learning rate with LDX treatment.

Inverse Softmax Temperature

Inverse softmax temperatures were relatively high across all time points. Typically, a larger inverse softmax temperature would indicate that participants were deterministically selecting the options with highest expected values (exploiting) as opposed to exploring. However, as the participants were not learning throughout the task, it is not necessary to comment on the participants' relative degree of exploration or exploitation.

Perseveration

Perseveration parameter values were low and negative. Normally, this would indicate that participants had less of a propensity toward repeating previously selected actions independently of their reward history (Gershman, 2020). However, as the participants were unable to learn the task reward contingencies, it is also irrelevant to report on this aspect of decision-making.

4.2 Possible Explanations for Unexpected Study Findings

It was expected that reinforcement learning, in some capacity, would have improved with BN symptoms. This was contrary to the study results which suggest that participant reinforcement learning did not increase with decreasing binge/purge behaviours. All participants received adequate task training prior to beginning the main task. Therefore, it was unlikely that participants could not optimize their performance due to an inability to appropriately navigate the task or understand the task state-transition structure.

One potential reason that the participants did not learn the task as their symptoms improved was because of insensitivity to reward. This idea is supported by current

literature which suggests that pathological binge eating in BN occurs as a result of reward hyposensitivity (Friederich et al., 2013). For example, individuals with BN may consume a large amount of appetitive food during a binge episode to stimulate their under-responsive reward system. BN participants with reward insensitivity would be less able to distinguish between the predicted and actual values of rewards, and this would limit their ability to learn the value of task actions/states (Nunes et al., 2018). Unfortunately, the decision-making paradigm used in our study cannot be used to detect sensitivity to reward. Therefore, no direct evidence exists to support this explanation.

It is also conceivable that participants did not effectively learn the task because they found the task rewards to be non-incentivizing. It may be necessary to measure participant reinforcement learning using disorder-specific reinforcers (or food reinforcers) instead of the reinforcers found on the two-step task (the money reward). The BN participants might have been relatively unmotivated by the money reward and impairment in learning and goal-directed control might be reduced if food-rewards are at stake. This topic is explored by other, recent studies, which show that generic monetary rewards are less motivating for psychiatric populations than they are for healthy controls (Wyckmans et al., 2019; Voon et al., 2015).

Third, it is possible that the participants did not learn with increasing LDX use, because the task did not appropriately measure participant reinforcement learning. This is unlikely, however, as healthy controls have been shown that they can learn on the novel two-step task by Kool et al. (2016) and can optimize their performance. We would expect

that the task could capture reinforcement learning in BN participants just as it has been used to measure reinforcement learning in non-clinical populations. This is especially true, given that our study task was identical to the novel task with the exception of our inclusion of a colour-blind friendly palette. We also showed that different reinforcement learning models, through simulated behaviour on the task, could be accurately identified statistically using our model fitting procedures. This suggests that the task does discriminate along various subcomponents of reinforcement learning.

Though it is unlikely that the task could not measure participant reinforcement learning, it is feasible that participant learning, and BN symptoms are independent of each other, or that BN pathology is related to some other factor not investigated by the task. Describing the exact causes of BN is a challenging process, because the neurobiology of the condition is poorly understood (Kaye, 2008). Due to a lack of understanding of BN pathogenesis, it is possible that the task didn't capture aspects of learning or cognition that are relevant to BN. Other neurobiological mechanisms that are implicated in the etiology and maintenance of binge eating include reduced executive control of attention ("cognitive interference control"), increased discounting of future-rewards ("delay-discounting"), impaired mental flexibility ("set-shifting") (Frank & Berner, 2020) and increased affective lability (Anestis et al., 2012). These aspects of behaviour were not measured by the study task, and therefore, no conclusions can be made regarding their influence on the participants' BN symptoms. It is also possible that participants experienced a reduction in binge episodes as a result of the appetite-suppressing effects of the study drug. Decreased appetite has been frequently reported among adults with

BED treated with LDX (McElroy et al., 2015). Thus, LDX may reduce pathologic over-eating in BN through appetite regulation, as opposed to increased learning.

4.3 Study Strengths

This is one of the first studies to examine the potential mechanisms by which LDX decreases binge/purge behaviour. In particular, the use of computational methods for this purpose offers a novel approach for assessing the interplay of neurobiology, symptom severity and treatment efficacy in BN.

The reinforcement learning task administered provided a computationally precise method of differentiating between participant model-based and model-free control (Kool et al., 2016). This task was developed by Kool et al. (2016) and is based on a version of the "two-step task" by Daw et al. (2011), which is reported as the dominant method of assessing these decision-making traits (Hasz & Redish, 2018). Additionally, the task improves the original Daw paradigm by incorporating a trade-off between decisional accuracy and computational demand (Kool et al., 2016). For these reasons, this particular reinforcement learning task was selected for use in our study.

An additional study strength was our implementation of quality control measures to ensure adequate usability and reliability of the task. Pilot testing verified that the task was able to be viewed in full resolution and was able to be run easily at the study location. Feedback from pilot testers also revealed the need to alter the original task colour scheme to incorporate a colour-blind friendly palette. This modified version of the task was used

to accommodate colour-blind individuals. Last, pilot testing confirmed that the task could be used to accurately detect the reinforcement learning parameters of interest.

4.4 Study limitations

The study has potential limitations. This is a single-arm, within-subject design across active treatment. There is no counterbalancing performance between drug and no drug conditions, nor is there a control group. As the study is exploratory in nature, the pilot data from the proposed research, will be used to inform larger studies with improved study design, for example, randomized controlled trials.

Second, the study included a narrow demographic of participants. Only female participants were recruited, and most participants identified as Caucasian. This precludes the generalization of study results to males, and other racial groups.

Third, the study did not examine the neurobiological changes that were associated with LDX treatment (i.e., through the use of neuroimaging techniques). Therefore, no conclusions can be drawn regarding the mechanism of action through which LDX produced its clinical effects. Further investigation of neural systems before and after LDX treatment is required to better understand neural dysfunction in BN, and how LDX acts on these systems to effectively reduce BN symptoms.

Fourth, the study used measures of binge eating and purging behaviour obtained through participant self-report data. There are a number of aspects of bias that accompany participant self-reporting including biases in memory, and social-desirability bias (Hebert et al., 1995), where participants report behaviours that they perceive to be more aligned with the intervention's goals. Such biases associated with self-report measures could complicate the interpretation of the study results.

Last, confounding factors may have contributed to the observed reduction in BN symptoms among the treated participants. For example, participants' use of dietary records, and their interactions with clinicians during in-person study visits, may have been partly responsible for their decrease in BN behaviour. While it may not be accurate to suggest that LDX treatment was wholly responsible for their reduction in BN pathology, the results are more applicable to a real-world setting, where multiple factors may contribute to the efficacy of a drug treatment.

4.5 Implications

Few studies have examined reward learning among individuals with BN. This preliminary investigation of decision-making in BN patients contributes to our existing understanding of the neurocognitive processes that are associated with binge/purge behaviour and provides insight into how LDX may function to reduce BN symptoms. It is possible that the participants' over-reliance on habitual action was one factor that contributed to the initial entrenchment of their maladaptive eating (Foerde et al., 2019).

This finding is of particular interest and may inspire future research that employs interventions to increase goal-directed behaviour among individuals with BN. Though it remains unclear how participants experienced a reduction in BN symptoms following LDX treatment without shifting to a primarily model-based approach, other studies would be able to explore alternative explanations for this finding.

Reinforcement learning tasks, such as the one used in our study, have the potential to improve clinical interventions. These tasks can be used to detect differences in learning or decision-making that may be implicated in certain psychiatric disorders. Despite this potential, our study identified some challenges that might prevent the task from assessing real-world BN behaviours. For example, the participants might not have been incentivized to learn the task because the task did not incorporate disorder-specific rewarding stimuli. The task should be altered to incorporate reinforcers that are relevant to BN (appetitive food rewards) to accurately capture participant decisional strategies. It is necessary that the task be personalized for use by BN patients before it can be used as a supportive tool in clinical decision-making.

Both the study and the research it inspires and may lead to an improved understanding of the neurological basis of BN through a mathematically informed, computational approach. This may improve pharmacological treatments for BN and provide a new perspective on brain-behaviour relationships in this psychiatric disorder.

4.6 Recommendations For Future Research

Literature suggests that reduced reward sensitivity may be of relevance in the genesis of BN behaviour (Friedrich et al., 2013). It is possible that participants were not able to learn the task as a result of this potential reward hyposensitivity. Therefore, it is recommended that future studies focus on testing reward sensitivity in participants with a task that is more specific to that. For example, the spatial orientation test by Derryberry and Reed (1994) has been used as a behavioural measure of reward sensitivity within the context of other eating disorder populations (Matton et al., 2017). This task measures differences in both attentional engagement and attentional disengagement associated with reward-related cues. The Sensitivity to Punishment/Sensitivity to Reward Questionnaire is another validated method for the assessment of reward sensitivity in eating disorder populations (Beck et al., 2009) and could also be used to detect reward sensitivity among individuals with BN.

It is also recommended that future studies examine participant reinforcement learning using both monetary and food-specific contexts. This is relevant because individuals with BN are posited to have altered responsiveness of the reward network to food stimuli compared with monetary rewards (Simon et al., 2016). Foerde et al. (2019) describe their protocol for comparing participant learning in anorexia nervosa patients using both food and monetary versions of a two-step task. In the food (illness relevant) task, participants were rewarded with food tokens that they could use to select from a variety of preferred food items at the end of each task. This use of monetary and disorder-specific reinforcers

would detect changes in participant decision-making that were related to differences in domain.

Future research investigating neurocognitive mechanisms of LDX pharmacotherapy in BN should also include neuroimaging. Functional brain imaging studies could be conducted to characterize the pharmacological effects of LDX on BN patient brain structure and function. In particular, these studies should investigate brain regions that are implicated in reward-based decision making, to deduce if pathologic binging and purging are related to dysfunction in these regions.

Last, replication of the study is recommended. This would reduce the likelihood that the observed experimental effects were caused merely by sampling variability.

4.7 Conclusions

Clinical psychiatry has long experienced a stagnation resulting from an overreliance on symptom-based definitions for mental disorders, without consideration of their neurobiological causes (Yahata et al., 2017). The recent application of computational psychiatry offers a potential solution to this problem, by quantitatively describing disorder-specific mechanisms. The results from our study will contribute to this emerging field by providing a novel description of BN-specific aberrations in decision-making. Uncovering the neurobiology of eating disorders such as BN, has the potential to re-conceptualize our understanding of these disorders, and to improve future methods for patient diagnosis and treatment.

REFERENCES

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology*, *42*(10), 1931-1939. https://doi.org/10.1038/npp.2017.108

Anestis, M. D., Peterson, C. B., Bardone-Cone, A. M., Klein, M. H., Mitchell, J. E., Crosby, R. D., Wonderlich, S. A., Crow, S. J., le Grange, D., & Joiner, T. E. (2009). Affective lability and impulsivity in a clinical sample of women with bulimia nervosa: the role of affect in severely dysregulated behavior. *International Journal of Eating Disorders*, *42*(3), 259-266. https://doi.org/10.1002/eat.20606

Anestis, M. D., Silva, C., Lavender, J. M., Crosby, R. D., Wonderlich, S. A., Engel, S. G., & Joiner, T. E. (2012). Predicting nonsuicidal self-injury episodes over a discrete period of time in a sample of women diagnosed with bulimia nervosa: An analysis of self-reported trait and ecological momentary assessment based affective lability and previous suicide attempts. *International Journal of Eating Disorders*, *45*(6), 808-811. https://doi.org/10.1002/eat.20947

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning, 47*(2), 235-256. https://doi.org/10.1023/A:1013689704352

Bäckman, L., Waris, O., Johansson, J., Andersson, M., Rinne, J. O., Alakurtti, K., Sovereign, A., Lainem, M., & Nyberg, L. (2017). Increased dopamine release after working-memory updating training: Neurochemical correlates of transfer. *Scientific Reports*, *7*(1), 1-10. http://doi.org/10.1038/s41598-017-07577-y

Balleine, B. W., & O' Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35*(1), 48-69. https://doi.org/10.1038/npp.2009.131

Baptista, T. (1999). Body weight gain induced by antipsychotic drugs: mechanisms and management. *Acta Psychiatrica Scandinavica*, *100*(1), 3-16. https://doi.org/10.1111/j.1600-0447.1999.tb10908.x

Barnett, R., Maruff, P., Vance, A., Luk, E.S.L., Costin, J., Wood, C., & Pantelis, C. (2001). Abnormal executive function in attention deficit hyperactivity disorder: The effect of stimulant medication and age on spatial working memory. *Psychological Medicine, 31*(6), 1107-15. http://doi.org/10.1017/S0033291701004172

Beck, I., Smits, D., Claes, L., Vandereycken, W., & Bijttebier, P. (2009). Psychometric evaluation of the behavioral inhibition/behavioral activation system scales and the sensitivity to punishment and sensitivity to reward questionnaire in a sample of eating disordered patients. *Personality and Individual Differences, 47*(5), 407-412. https://doi.org/10.1016/j.paid.2009.04.007

Bellman, R., & Rand Corporation. (1957). *Dynamic Programming*. Princeton University Press.

Berner, L. A., & Marsh, R. (2014). Frontostriatal circuits and the development of bulimia nervosa. *Frontiers in Behavioral Neuroscience*, *8*, 395. https://doi.org/10.3389/fnbeh.2014.00395

Boulay, L., Labelle, A., Bourget, D., Robertson, S., Habib, R., Tessier, P., Tombaugh, T., Milin, R. (2007). Dissociating medication effects from learning and practice effects in a neurocognitive study of schizophrenia: Olanzapine versus haloperidol. *Cognitive Neuropsychiatry, 12*(4), 322-338. https://doi.org/10.1080/13546800601069534

Broft, A. I., Berner, L. A., Martinez, D., & Walsh, B. T. (2011). Bulimia nervosa and evidence for striatal dopamine dysregulation: a conceptual review. *Physiology & Behavior*, *104*(1), 122-127. https://doi.org/10.1016/j.physbeh.2011.04.028

Brown, T. E., Brams, M., Gao, J., Gasior, M., & Childress, A. (2010). Open-Label Administration of Lisdexamfetamine Dimesylate Improves Executive Function Impairments and Symptoms of Attention-Deficit/Hyperactivity Disorder in Adults. *Postgraduate Medicine*, *122*(5), 7-17. http://doi.org/10.3810/pgm.2010.09.2196

Brown, R. L., Leonard, T., Saunders, L. A., & Papasouliotis, O. (2001). A two-item conjoint screen for alcohol and other drug problems. *Journal of the American Board of Family Practice*, *14*(2), 95-106. https://www.jabfm.org/content/jabfp/14/2/95.full.pdf

Calugi, S., Ricca, V., Castellini, G., Lo Sauro, C., Ruocco, A., Chignola, E., El Ghoch, M., & Dalle Grave, R. (2015). The eating disorder examination: reliability and validity of the Italian version. *Eating and Weight Disorders, 20*(4), 505-511. https://doi.org/10.1007/s40519-015-0191-2

Childress, A. C., Arnold, V., Adeyi, B., Dirks, B., Babcock, T., Scheckner, B., Lasser, R., & Lopez, F. A. (2014). The effects of lisdexamfetamine dimesylate on emotional lability in children 6 to 12 years of age with ADHD in a double-blind placebo-controlled trial. *Journal of Attention Disorders*, *18*(2), 123-132. https://doi.org/10.1177/1087054712448252

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature, 482*, 85-88. https://doi.org/10.1038/nature10754

Conway, A., Kane, R., Bunting, A., Hambrick, M., Wilhelm, J., & Engle, F. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786. https://doi.org/10.3758/BF03196772

da Silva, C. F., & Hare, T. A. (2019). Model-free or muddled models in the two-stage task?. *bioRxiv*, 682922. https://doi.org/10.1101/682922

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204-1215. https://doi.org/10.1016/j.neuron.2011.02.027

Derryberry, D., & Reed, M. A. (1994). Temperament and attention: orienting toward and away from positive and negative signals. *Journal of Personality and Social Psychology*, *66*(6), 1128-1139. https://doi.org/10.1037//0022-3514.66.6.1128

De Wit, S., Standing, H. R., DeVito, E. E., Robinson, O. J., Ridderinkhof, K. R., Robbins, T. W., & Sahakian, B. J. (2012). Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology*, *219*(2), 621-631. https://doi.org/10.1007/s00213-011-2563-2

Diederen, K. M., Ziauddeen, H., Vestergaard, M. D., Spencer, T., Schultz, W., & Fletcher, P. C. (2017). Dopamine Modulates Adaptive Prediction Error Coding in the Human Midbrain and Striatum. *Journal of Neuroscience*, *37*(7), 1708-1720. https://doi.org/10.1523/JNEUROSCI.1979-16.2016

Dynesen, A. W., Bardow, A., Petersson, B., Nielsen, L. R., & Nauntofte, B. (2008). Salivary changes and dental erosion in bulimia nervosa. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*, *106*(5), 696-707. https://doi.org/10.1016/j.tripleo.2008.07.003

Foerde, K., Daw, N., Rufin, T., Walsh, B. T., Shohamy, D., & Steinglass, J. (2019). Deficient goal-directed control in a population characterized by extreme goal pursuit. *MedRxiv*. https://doi.org/10.1101/19002089

Frank, G. K., & Berner, L. A. (Eds.). (2020). *Binge Eating: A Transdiagnostic Psychopathology*. Springer Nature.

Frank, M., Doll, B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience, 12*(8), 1062-108. https://doi.org/10.1038/nn.2342

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311-16316. https://doi.org/10.1073/pnas.0706111104

Frank, G., Reynolds, J., Shott, M., & O'Reilly, R. (2011). Altered temporal difference learning in bulimia nervosa. *Biological Psychiatry*, *70*(8), 728-735. http://doi.org/10.1016/j.biopsych.2011.05.011

Friederich, H. C., Wu, M., Simon, J. J., & Herzog, W. (2013). Neurocircuit function in eating disorders. *International Journal of Eating Disorders*, *46*(5), 425-432. http://doi.org/10.1002/eat.22099

Furlong, T. M., Jayaweera, H. K., Balleine, B. W., & Corbit, L. H. (2014). Binge-like consumption of a palatable food accelerates habitual control of behavior and is dependent on activation of the dorsolateral striatum. *Journal of Neuroscience*, *34*(14), 5012-5022. http://doi.org/10.1523/JNEUROSCI.3707-13.2014

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, *204*, 104394. https://doi.org/10.1016/j.cognition.2020.104394

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *ELife*, *5*, e11305. https://doi.org/10.7554/eLife.11305

Griffiths, K. R., Yang, J., Touyz, S. W., Hay, P. J., Clarke, S. D., Korgaonkar, M. S., Gomes, L., Foster, S., & Kohn, M. R. (2019). Understanding the neural mechanisms of lisdexamfetamine dimesylate (LDX) pharmacotherapy in Binge Eating Disorder (BED): a study protocol. *Journal of Eating Disorders*, *7*(1), 1-10. https://doi.org/10.1186/s40337-019-0253-3

Grob, S., Pizzagalli, D. A., Dutra, S. J., Stern, J., Mörgeli, H., Milos, G., Schnyder, U., & Hasler, G. (2012). Dopamine-related deficit in reward learning after catecholamine depletion in unmedicated, remitted subjects with bulimia nervosa. *Neuropsychopharmacology*, *37*(8), 1945-1952. https://doi.org/10.1038/npp.2012.41

Grob, S., Stern, J., Gamper, L., Moergeli, H., Milos, G., Schnyder, U., & Hasler, G. (2015). Behavioral responses to catecholamine depletion in unmedicated, remitted subjects with bulimia nervosa and healthy subjects. *Biological Psychiatry*, *77*(7), 661-667. https://doi.org/10.1016/j.biopsych.2013.09.013

Haedt-Matt, A. A., & Keel, P. K. (2011). Revisiting the affect regulation model of binge eating: a meta-analysis of studies using ecological momentary assessment. *Psychological Bulletin, 137*(4), 660-681. https://doi.org/10.1037/a0023660

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. https://doi.org/10.1016/j.jpsychires.2021.01.046

Hasz, B., & Redish, A. D. (2018). Behavioral Correlates of Model-based and Model-free Uncertainty are Revealed by a Two-step Decision Task for Rats. *Frontiers in Integrative Neuroscience*, *12*, 30. https://doi.org/10.3389/fnint.2018.00030

Hebert, J. R., Clemow, L., Pbert, L., Ockene, I. S., & Ockene, J. K. (1995). Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *International Journal of Epidemiology, 24*(2), 389-398. https://doi.org/10.1093/ije/24.2.389

Homan, P., Grob, S., Milos, G., Schnyder, U., Eckert, A., Lang, U., & Hasler, G. (2015). The role of BDNF, leptin, and catecholamines in reward learning in bulimia nervosa. *International Journal of Neuropsychopharmacology*, *18*(5), 1-8. https://doi.org/10.1093/ijnp/pyu092

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Computational Biology*, *7*(4). https://doi.org/10.1371/journal.pcbi.1002028

Jimerson, D., Lesem, M., Kaye, W., & Brewerton, T. (1992). Low serotonin and dopamine metabolite concentrations in cerebrospinal fluid from bulimic patients with frequent binge episodes. *Archives of General Psychiatry, 49*(2), 132-138. http://doi.org/10.1001/archpsyc.1992.01820020052007

Jocham, G., Klein, T. A., & Ullsperger, M. (2014). Differential modulation of reinforcement learning by D2 dopamine and NMDA glutamate receptor antagonism. *Journal of Neuroscience, 34*(39), 13151-13162. https://doi.org/10.1523/JNEUROSCI.0757-14.2014

Kaye, W. (2008). Neurobiology of anorexia and bulimia nervosa. *Physiology & Behavior*, *94*(1), 121-135. http://doi.org/10.1016/j.physbeh.2007.11.037

Kayser, A., Mitchell, J., Weinstein, D., & Frank, M. (2015). Dopamine, Locus of Control, and the Exploration-Exploitation Tradeoff. *Neuropsychopharmacology*, *40*(2), 454-462. http://doi.org/10.1038/npp.2014.193

Kekic, M., Bartholdy, S., Cheng, J., McClelland, J., Boysen, E., Musiat, P., O'Daly, O., Campbell, I., & Schmidt, U. (2016). Increased temporal discounting in bulimia nervosa. *International Journal of Eating Disorders*, *49*(12), 1077-1081. http://doi.org/10.1002/eat.22571

Keshen, A. R., Dixon, L., Ali, S. I., Helson, T., Nunes, A., Milliken, H., Gamberg, S., Sadek, J., Kaplan, A., & McElroy, S. L. (2021). A feasibility study evaluating lisdexamfetamine dimesylate for the treatment of adults with bulimia nervosa. *International Journal of Eating Disorders*, *54*(5), 872-878. https://doi.org/10.1002/eat.23480

Keshen, A. R., & Helson, T. (2017). Preliminary evidence for the off-label treatment of bulimia nervosa with psychostimulants: six case reports. *Journal of Clinical Pharmacology.* Advance online publication. http://doi.org//10.1002/jcph.868

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine, 35*(02), 245-256. http://doi.org/10.1017/s0033291704002892

Kessler, R. C., Akiskal, H. S., Angst, J., Guyer, M., Hirschfeld, R. M., Merikangas, K. R., & Stang, P. E. (2006). Validity of the assessment of bipolar spectrum disorders in the WHO CIDI 3.0. *Journal of Affective Disorders*, *96*(3), 259-269. http:/doi.org/10.1016/j.jad.2006.08.018

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments & Computers, 31*(3), 429-432. http:/doi.org/10.3758/bf03200722

Kool, W., Cushman, F., & Gershman, S. (2016). When Does Model-Based Control Pay Off? *PLOS Computational Biology, 12*(8), E1005090. http://doi.org/10.1371/journal.pcbi.1005090

Matton, A., de Jong, P., Goossens, L., Jonker, N., Van Malderen, E., Vervaet, M., De Schryver, N., & Braet, C. (2017). Sensitivity for Cues Predicting Reward and Punishment in Young Women with Eating Disorders. *European Eating Disorders Review*, *25*(6), 501-511. https://doi.org/10.1002/erv.2541

McElroy, S. L., Hudson, J. I., Mitchell, J. E., Wilfley, D., Ferreira-Cornwell, M. C., Gao, J., Wang, J., Whitaker, T., Jonas, J., Gasior, M. (2015). Efficacy and safety of lisdexamfetamine for treatment of adults with moderate to severe binge-eating disorder: A randomized clinical trial. *JAMA Psychiatry, 72*(3), 235-246. http://doi.org/10.1001/jamapsychiatry.2014.2162

Mehler, P. S. (2011). Medical complications of bulimia nervosa and their treatments. *International Journal of Eating Disorders*, *44*(2), 95-104. https://doi.org/10.1002/eat.20825

Mehler, P. S., & Rylander, M. (2015). Bulimia Nervosa - medical complications. J*ournal of Eating Disorders*, *3*, 12. https://doi.org/10.1186/s40337-015-0044-4

Michaelides, M., Thanos, P. K., Volkow, N. D., & Wang, G. J. (2012). Dopamine-related frontostriatal abnormalities in obesity and binge-eating disorder: emerging evidence for developmental psychopathology. *International Review of Psychiatry*, *24*(3), 211-218. https://doi.org/10.3109/09540261.2012.679918

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience, 16*(5), 1936-1947. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996

Morris, L., Baek, K., Kundu, P., Harrison, N., Frank, M., & Voon, V. (2015). Biases in the Explore-Exploit Tradeoff in Addictions: The Role of Avoidance of Uncertainty. *Neuropsychopharmacology*, *41*(4), 940-948. https://doi.org/10.1038/npp.2015.208

Nunes, A., Helson, T. J., Dixon, L. M., Trappenberg, T. P., & Keshen, A. R. (2018). Characteristic impairments of goal-directed and habitual control in bulimia nervosa. *bioRxiv*, 348250. https://doi.org/10.1101/348250

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, Reward, and Decision Making. *Annual Review of Psychology, 68*(1), 73-100. https://doi.org/10.1146/annurev-psych-010416-044216

Onysk, J., & Series, P. (2020). The effect of body image preoccupation on decision making in eating disorders. *medRxiv*. https://doi.org/10.1101/2020.11.15.20232090

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The Curse of Planning: Dissecting Multiple Reinforcement-Learning Systems by Taxing the Central Executive. *Psychological Science*, *24*(5), 751-761. https://doi.org/10.1177/0956797612463080

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(52), 20941-20946. https://doi.org/10.1073/pnas.1312011110

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Reiter, A., Heinze, H., Schlagenhauf, F., & Deserno, L. (2016). Impaired Flexible Reward-Based Decision-Making in Binge Eating Disorder: Evidence from Computational Modeling and Functional Neuroimaging. *Neuropsychopharmacology*, *42(3)*, 1-10. http://doi.org/10.1038/npp.2016.95

Rigoux, L., Stephan, K., Friston, K., & Daunizeau, J. (2014). Bayesian model selection for group studies - Revisited. *NeuroImage*, *84*, 971-85. http://doi.org/10.1016/j.neuroimage.2013.08.065

Rushing, J., Jones, L., & Carney, C. (2003). Bulimia Nervosa: A Primary Care Review. *Primary Care Companion to the Journal of Clinical Psychiatry*, *5*(5), 217-224. https://doi.org/10.4088/pcc.v05n0505

Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *The Journal of Neuroscience*, *29*(48), 15104-15114. https://doi.org/10.1523/JNEUROSCI.3524-09.2009

Salmi, J., Ritakallio, L., Fellman, D., Ellfolk, U., Rinne, J. O., & Laine, M. (2020). Disentangling the role of working memory in Parkinson's disease. *Frontiers in Aging Neuroscience*, *12*, 572037. https://doi.org/10.3389/fnagi.2020.572037

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599. http://doi.org/10.1126/science.275.5306.1593

Sharp, M. E., Foerde, K., Daw, N. D., & Shohamy, D. (2016). Dopamine selectively remediates 'model-based' reward learning: a computational approach. *Brain*, *139*(2), 355-364. http://doi.org/10.1093/brain/awv347

Simon, J. J., Skunde, M., Walther, S., Bendszus, M., Herzog, W., & Friederich, H. C. (2016). Neural signature of food reward processing in bulimic-type eating disorders. *Social Cognitive and Affective Neuroscience*, *11*(9), 1393-1401. https://doi.org/10.1093/scan/nsw049

Skunde, M., Walther, S., Simon, J. J., Wu, M., Bendszus, M., Herzog, W., & Friederich, H. C. (2016). Neural signature of behavioural inhibition in women with bulimia nervosa. *Journal of Psychiatry & Neuroscience*, *41*(5), E69-E78. https://doi.org/10.1503/jpn.150335

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., & Engel, S. G. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology*, *75*(4), 629-638. https://doi.org/10.1037/0022-006X.75.4.629

Steinglass, J. E., & Walsh, B. T. (2016). Neurobiological model of the persistence of anorexia nervosa. *Journal of Eating Disorders*, *4*, 19. https://doi.org/10.1186/s40337-016-0106-2

Stice, E., Telch, C. F., & Rizvi, S. L. (2000). Development and validation of the eating disorder diagnostic scale: A brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychological Assessment*, *12*(2), 123-131. https://doi.org/10.1037//1040-3590.12.2.123

Stine, S. M., Krystal, J. H., Petrakis, I. L., Jatlow, P. I., Heninger, G. R., Kosten, T. R., & Charney, D. S. (1997). Effect of alpha-methyl-para-tyrosine on response to cocaine challenge. *Biological Psychiatry*, *42*(3), 181-190. https://doi.org/10.1016/s0006-3223(96)00331-9

Sutton, R. S. (1998) Learning to Predict by the Methods of Temporal Differences. *Machine Learning*, *3*(1), 9-44. http://doi.org/10.1007/BF00115009

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychological Review*, *88*(2), 135. https://doi.org/10.1037/0033-295X.88.2.135

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: an introduction. MIT Press.

Sutton, R. S., McAllester, D. A., Singh, S., & Mansour, Y. (1999). Policy Gradient Methods for Reinforcement Learning with Function Approximation. *Neural Information Processing Systems, 12,* 1057-1063. https://proceedings.neurips.cc/paper/1999/file/464d828b85b0bed98e80ade0a5c43b0f-Paper.pdf

Thaler, L., Groleau, P., Badawi, G., Sycz, L., Zeramdini, N., Too, A., Israel, M., Joober, R., Bruce, K. R., & Steiger, H. (2012). Epistatic interactions implicating dopaminergic genes in bulimia nervosa (BN): Relationships to eating- and personality-related psychopathology. *Progress in Neuropsychopharmacology & Biological Psychiatry, 39*(1), 1201-28.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements, 2*(4), i–109. https://doi.org/10.1037/h0092987

Volkow, N. D., Fowler, J. S., Wang, G. J., Baler, R., & Telang, F. (2009). Imaging dopamine's role in drug abuse and addiction. *Neuropharmacology*, *56*, 3-8. https://doi.org/10.1016/j.neuropharm.2008.05.022

von der Malsburg, T. (2015). Py-Span-Task – A Software for Testing Working Memory Span. Zenodo. http://doi.org/10.5281/zenodo.18238

Voon, V. (2015). Cognitive biases in binge eating disorder: the hijacking of decision making. *CNS Spectrums*, *20*(6), 566-573. https://doi.org/10.1017/S1092852915000681

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., & Bullmore, E. T. (2015). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*, 345-352. https://doi.org/10.1038/mp.2014.44

Wagner, A., Aizenstein, H., Venkatraman, V. K., Bischoff-Grethe, A., Fudge, J., May, J. C., Frank, G. K., Bailer, U. F., Fischer, L., Putnam, K., & Kaye, W. H. (2010). Altered striatal response to reward in bulimia nervosa after recovery. *International Journal of Eating Disorders*, *43*(4), 289-294. http://doi.org/10.1002/eat.20699

Wang, T., Wu, X., Jiang, J., Liu, C., & Zhu, M. (2020). Functional neural interactions during adaptive reward learning: A functional magnetic resonance imaging study. *International Journal of Imaging Systems and Technology*, *30*(1), 92-103. https://doi.org/10.1002/ima.22387

Watson, H. J., McLagan, N., Zerwas, S. C., Crosby, R. D., Levine, M. D., Runfola, C. D., Peat, C. M., Moessner, M., Zimmer, B., Hofmeier, S. M., Hamer, R. M., Marcus, M. D., Bulik, C. M., & Crow, S. J. (2018). Cost-Effectiveness of Internet-Based Cognitive-Behavioral Treatment for Bulimia Nervosa: Results of a Randomized Controlled Trial. *Journal of Clinical Psychiatry*, *79*(1), 16m11314. https://doi.org/10.4088/JCP.16m11314

Weiss, N. A. (2015). wPerm: Permutation Tests. R package version 1.0.1. https://CRAN.R-project.org/package=wPerm

Wigal, T., Brams, M., Gasior, M., Gao, J., & Giblin, J. (2011). Effect size of lisdexamfetamine dimesylate in adults with attention-deficit/hyperactivity disorder. *Postgraduate Medicine*, *123*(2), 169-176. https://doi.org/10.3810/pgm.2011.03.2275

Wilson, G. T., Grilo, C. M., & Vitousek, K. M. (2007). Psychological treatment of eating disorders. *American Psychologist*, *62*(3), 199. https://doi.org/10.1037/0003-066X.62.3.199

Wong, C. G., & Stevens, M. C. (2012). The effects of stimulant medication on working memory functional connectivity in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *71*(5), 458-466. https://doi.org/10.1016/j.biopsych.2011.11.011

Wu, M., Brockmeyer, T., Hartmann, M., Skunde, M., Herzog, W., & Friederich, H. (2016). Reward-related decision making in eating and weight disorders: A systematic review and meta-analysis of the evidence from neuropsychological studies. *Neuroscience and Behavioural Reviews*, *61*, 177-196. https://doi.org/10.1016/j.neubiorev.2015.11.017

Wunderlich, K., Smittenaar, P. & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron*, *75*, 418-424. https://doi.org/10.1016/j.neuron.2012.03.042

Wyckmans, F., Otto, A. R., Sebold, M., Daw, N., Bechara, A., Saeremans, M., Kornreich, C., Chatard, A., Jaafari, N., & Noël, X. (2019). Reduced model-based decision-making in gambling disorder. *Scientific Reports*, *9*(1), 19625. https://doi.org/10.1038/s41598-019-56161-z

Yahata, N., Kasai, K., & Kawato, M. (2017). Computational neuroscience approach to biomarkers and treatments for mental disorders. *Psychiatry and Clinical Neurosciences*, *71*(4), 215-237. https://doi.org/10.1111/pcn.12502

Yu, J., & Selby, E. A. (2013). The interaction between affective lability and interpersonal problems in binge eating. *Journal of Social and Clinical Psychology*, *32*(5), 465-481. https://doi.org/10.1521/jscp.2013.32.5.465

APPENDIX A Pre-Screening Questions/Measures

Items screened	Description of items
Contact Information	Full Name Email & Phone number
Demographics	Sex Year/month of birth Current or intended pregnancy (within the next year) Current breastfeeding status Current stimulant medication Current antidepressant medication Recent dose change of any psychiatric medication (past 4-weeks) Current psychotherapy for BN Previous eating disorder treatment Recreational use of illicit substances in the previous 2-months Syrup of ipecac use History of anorexia nervosa Allergy to amphetamines or sensitivity to stimulants Diagnosis of glaucoma Diagnosis of hyperthyroidism Personal or family history of cardiovascular disease (specify)
Eating Disorder Diagnostic Scale (EDDS; Stice et al., 2000)	EDDS is a validated screening measure for eating disorder symptoms and preliminary eating disorder diagnoses. Height and current, lowest and highest weight are included.

Items screened	Description of items
ADHD Diagnosis & ASRS (Adult ADHD Self-Report Scale; Kessler et al., 2005)	Previous diagnosis of ADHD The ASRS is a brief, validated screening measure for ADHD.
Substance Use/Cut Down (Brown et al., 2001)	A 2-item conjoint screen for alcohol or drug problems
Psychotic Disorder Diagnosis	Previous diagnosis of a psychotic disorder
Bipolar Diagnosis & Bipolar WHO-CIDI (Kessler et al., 2006)	Previous diagnosis of bipolar disorder Bipolar WHO-CIDI is a brief, validated screening measure for bipolar disorder.

Note. EDDS= Eating Disorder Diagnostic Scale; ADHD= Attention deficit hyperactivity disorder; ASRS= Adult ADHD Self-Report Scale; WHO-CIDI= World Health Organization Composite International Diagnostic Interview.

APPENDIX B Screening Visit

Screening Procedure	Description of Screening Procedure
1. Consent Discussion	Evaluation of participant competence Participant consent discussed/ received
2. Medical Assessment	Medical history Vital signs (height, weight, blood pressure, heart rate) EKG Blood draw Blood Profile & Chemistry, Drug Screen (At discretion of principal investigator), Pregnancy Screen (if applicable) Participant eligibility reaffirmed
3. Psychological Assessment	Structured Clinical Interview for DSM 5 Disorders (Research Version) (SCID-5-RV) to screen for comorbid disorders (see Eligibility criteria) Columbia Suicide Severity Rating Scale (C-SSRS)(Screening Version) to screen for suicide risk

All participants were medically/ psychologically assessed by the principal investigator. As reflected by the inclusion/exclusion criteria (Appendix C), the study recruited participants who were both a) experiencing moderate to extreme degrees of pathology, and b) at a lower risk of abusing the appetite suppressing effects of LDX.

APPENDIX C Criteria for Participant Inclusion/Exclusion

Inclusion Criteria:

• 18-55 years of age and signed consent

• Moderate to extreme BN (SCID-5-RV)

• BMI between 21-30 kg/m^2

• Ability to swallow capsule consistently as per self-report

• Females who are not breast-feeding and who are not of childbearing potential (last menstruation at least 24 months prior to baseline, or have undergone tubal ligation, or have undergone hysterectomy)

• Females of childbearing potential who have a negative serum pregnancy test prior to study enrollment and who agree to use a reliable method of birth control (oral contraceptives, contraceptive injections, contraceptive patch, intrauterine device, partner with vasectomy, abstinence or barrier methods such as condoms, vaginal diaphragm with spermicide or sponge) during the study and for one month following the last dose of the study drug

Exclusion Criteria:

Participants were excluded from the study if they met any one of the following criteria:

• Co-morbid bipolar disorder, psychotic disorder, moderate-severe depression, and/or ADHD according to SCID-5-RV

• Previous history of anorexia nervosa (due to elevated risk of problematic weight loss secondary to stimulant use)

• Severely restricting eating behaviours defined as routinely (>2 days per week) eating less than two meals a day, or at the investigator's discretion

• Clinically meaningful abnormalities in laboratory tests or electrocardiography results as determined by the principal investigator

• Personal or family history of cardiovascular disease that could increase vulnerability to the sympathomimetic effects of stimulants (e.g., structural cardiac abnormalities, cardiomyopathy, serious heart arrhythmia, advanced arteriosclerosis, or coronary artery disease) or any current symptomatic cardiovascular disease, as determined by the principal investigator, and/or in consultation with cardiologist (as needed)

• Participant has moderate to severe hypertension (>140/90 mmHg)

• Participant is receiving psychotherapy deemed by the investigators to be specifically treating BN

• Participant is taking or has taken a stimulant within the past 3 months

• Participant is on another psychotropic medication, and the dose has been changed 4 weeks prior to study drug initiation or the participant is on an antipsychotic medication

• History of substance use disorder in the preceding 6 months (or more distant at supervisor discretion) or a lifetime history of stimulant substance use disorder

• Participant is taking or has taken a monoamine oxidase inhibitor (MAOI) within the last 14 days

• Participant is pregnant, plans to become pregnant, or is nursing

• Participant uses syrup of ipecac (to self-induce vomiting)

• Participant is considered a suicide risk, according to the C-SSRS (Screening Version), and at the discretion of the principal investigator

• Participant has a known allergy to amphetamines, or other non-medical ingredients in LDX, or is sensitive to, is allergic to, or has had a reaction to other stimulant medications

• Participant has been diagnosed with glaucoma

• Participant has been diagnosed with hyperthyroidism

• Participant has insufficient knowledge of English language

APPENDIX D Patient Safety Monitoring

Participants were instructed to report all treatment emergent adverse events (TEAEs) to the principal investigator. Participants were instructed to visit the ER in the event of a psychiatric or medical emergency. Non-emergent questions regarding the study were directed toward the principal investigator or other study personnel.

All serious or unexpected adverse reactions were reported to the Nova Scotia Health Authority Research Ethics Board and Health Canada by the principal investigator. Other adverse effects were monitored and recorded by study investigators. Participants found to have met any of the criteria for discontinuation were automatically withdrawn from the study.

Study Discontinuation Criteria:

• The participant's BMI falls below 20, or they experience a rapid weight reduction (≥5% of body weight in a given month), or they become excessively restrictive in their eating patterns (defined as routinely [> 2 days a week] eats less than 2 meals a day), or are otherwise suspected of abusing LDX's appetite suppressing effects

• Participant's non-study medications change during the study period in such a way that could interfere with study outcomes

• Participant violates protocol, withdraws from study, or experiences a serious adverse event

• Participant is suspected of misusing study drug or other substances (i.e. as indicated by urine drug screen)

• Participant experiences suicidal ideation and is considered a suicide risk, or attempts suicide during treatment

• Participant becomes pregnant

• Development of laboratory, EKG, or vital sign abnormalities deemed by the principal investigator to be medically concerning (i.e., potassium, chloride or sodium abnormalities, hypoglycemia, prolonged QTc, hypertension, and tachycardia)

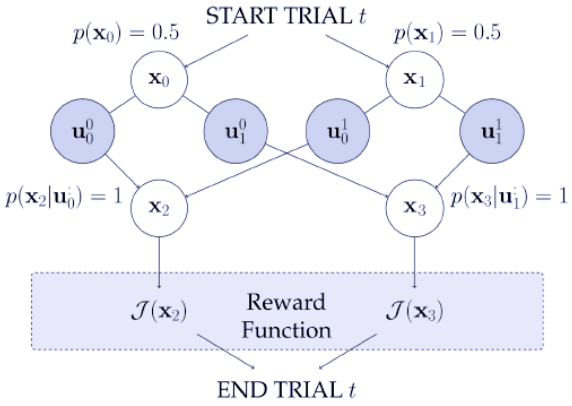
• New information shows that the study is not in the participant's best interests

APPENDIX E Patient Schedule

			Titration Phase				Maintenance Phase					Follow-Up
	Pre-Screen	Screen-ing	W1	W2	W3	W4	W5	W6	W7	W8	W9 Post	W10
In-Person Study Visit		X	X	X	X	X	X		X		X	X
LDX Dosage (mg/day)		__	30	50	70	50 or 70	50 or 70	50 or 70	50 or 70	50 or 70	__	__
Pre-Screening												
Online Pre-Screen	X											
Telephone Interview	X											
Screening/ Scheduled Procedures												
Consent Discussion		X										
Screening Interview- (SCID-5 + Eligibility)		X										
Blood Profile		X										
Drug Screen		X										
Pregnancy Screen		X										
EKG		X									X	
Blood Work		X					X				X	
Vital Signs/ Weight		X	X	X	X	X	X		X		X	X
Suicidality Screen			X	X	X	X	X	X	X	X	X	X
Adverse Events				X	X	X	X	X	X	X	X	X

			Titration Phase				Maintenance Phase					Follow-Up
Computer-Based Tasks												
2-Step Task			X	X							X	X
OSPAN			X	X							X	X
Clinician-Administered Measures												
EDE 17.0D			X								X	
Self-Report Measures												
Dietary Records			X	X	X	X	X	X	X	X	X	X

APPENDIX F Contextual Bandit Reinforcement Task

Block (and their code names)	Description
ID Collection	Participant's study ID and Date are input and held as a global variable to which other data from the task are linked.
"Real Trials"	 Individuals start trial *t* by entering either state x_0 (state with spaceships A and B) or state x_1 (state with spaceships C and D). If participant starts in state x_0, they may choose either action u_0^0 (spaceship A) or u_1^0 (spaceship B). If participant starts in state x_1, they can choose either action u_0^1 (spaceship C) or u_1^1 (spaceship D). At each trial, spaceship pairs are placed on either the left or right of the screen according to a probability of 0.5. The choice of spaceship leads deterministically to state x_2 (one of the planets) or state x_3(the other planet) where the participant receives a reward if he or she presses the "space bar" key on the keyboard. Between trials, participants are presented with a fixation cross in order to retain attention to the centre of the screen.

APPENDIX G Test-Retest Reliability of Covariate Measures

Covariate Measure	Test-retest correlations
Eating Disorder Examination 17.0 D (EDE 17.0D)	In a previous study by Calugi et al. (2015) EDE test-retest reliability (2-23 day retest period) was good to excellent for objective binge episodes/days (r=0.99), vomiting episodes (r=0.94), laxative use episodes (r=0.92), diuretic use episodes (r=1.00), and excessive exercise days (r=0.82), but was unsatisfactory for subjective binge days/episodes (r=0.36).
Operation Span Task (OSPAN)	A previous study by Klein and Fiss (1999) indicated that test-retest correlations for the OSPAN were sufficient at 3-week (r=0.73) 7-week (r=0.81), and 10-week (r=0.67) intervals between test administrations.

APPENDIX H Reinforcement Learning Models

Model	Component	Equation	Parameters
Hybrid	Policy	Softmax	β: Inverse softmax temperature (choice consistency)
	Value function	Model-free (Q^{MF}): SARSA Model-based (Q^{MB}): Bellman Integration: $Q = \omega Q^{MB} + (1-\omega)Q^{MF}$	Model free: ● Learning rate: $0 \leq \alpha \leq 1$ Integration: ● Model-based/model-free balance: $0 \leq \omega \leq 1$
Model-free	Hybrid model with ω fixed to 0		
Model-based	Hybrid model with ω fixed to 1		

APPENDIX I Results from Paired Permutation Testing

.

Population 1	Population 2	Mean Difference	H_0	H_a	*p*
OBE (T3)	OBE (T1)	-35.11	Identical	Shifted left	< 0.001
OBD (T3)	OBD (T1)	-16.84	Identical	Shifted left	< 0.001
SBE (T3)	SBE (T1)	-3.579	Identical	Shifted left	0.0064
SBD (T3)	SBD (T1)	-3.105	Identical	Shifted left	0.0059
CB (T3)	CB (T1)	-38.85	Identical	Shifted left	< 0.001
RR (T3)	RR(T1)	0.1511	Identical	Shifted right	0.0702
RR (T3)	RR (T4)	0.0586	Identical	Shifted right	0.313
RR (T1)	RR (T4)	-0.0925	Identical	Shifted	0.535
OSPAN (T3)	OSPAN (T1)	-0.9286	Identical	Shifted	0.887

Population 1	Population 2	Mean Difference	H_0	H_a	*p*
OSPAN (T3)	OSPAN (T4)	-8.500	Identical	Shifted	0.0138
OSPAN (T1)	OSPAN (T4)	-7.571	Identical	Shifted	0.223

Note. Results of permutation paired location test based on 9999 replications; OBE= objective binge episodes; OBD= objective binge days; SBE= subjective binge episodes; SBD= subjective binge days; CB= compensatory behaviours (past 28 days); RR= reward rate; OSPAN= Operation span task score; Shifted left= The distribution of the variable on the first population has systematically smaller values than that of the variable on the second population; Shifted right= The distribution of the variable on the first population has systematically larger values than that of the variable on the second population; Shifted= The distribution of the variable on the first population has either systematically smaller values or systematically larger values than that of the variable on the second population.

APPENDIX J Estimation of Model Parameters (Model-Free Agent with Learning Rate, Inverse Softmax Temperature and Perseveration)

Participant ID	Time	α	β	ρ
A	T1	0.0018	45.06	-0.70552
A	T3	0.0140	0.0104	-0.1991
A	T4	0.0014	88.43	-0.2239
E	T1	0.0012	46.81	-0.3450
E	T3	0.0028	41.42	-0.4575
E	T4	0.0106	9.295	-0.3933
F	T1	0.0203	1.086	0.9299
F	T3	0.0010	105.0	-0.2246
F	T4	0.0017	81.10	-0.6789
G	T1	0.0029	50.29	-0.1660
G	T3	0.0026	36.79	-0.0627
G	T4	0.0016	89.94	-0.4221
H	T1	0.0021	68.50	-0.4953
H	T3	0.0051	11.014	-0.3683
H	T4	0.0027	80.06	-0.2336
I	T1	0.0049	30.68	-0.0593
I	T3	0.0078	48.95	-0.5278
I	T4	0.0043	56.43	-0.7028
J	T1	0.0337	0.5628	0.1008
J	T3	0.0145	8.044	-0.5059

Participant ID	Time	α	β	ρ
J	T4	0.0008	99.30	-0.1692
M	T1	0.0013	57.02	-0.3713
M	T3	0.0415	0.6696	-0.0420
M	T4	0.0039	56.73	-0.3585
N	T1	0.0057	34.58	-0.2432
N	T3	0.0011	68.50	-0.5610
N	T4	0.0007	98.75	-0.3042
O	T1	0.0010	42.41	0.0984
O	T3	0.0059	42.29	-0.4663
O	T4	0.0003	115.6	-0.0308
P	T1	0.0021	45.49	-0.6029
P	T3	0.0019	70.71	-0.5259
P	T4	0.0016	92.27	-0.7287
Q	T1	0.0014	45.26	-0.3082
Q	T3	0.0035	16.28	-0.4172
Q	T4	0.0026	88.98	-0.7585
S	T1	0.0077	18.36	0.0014
S	T3	0.0056	22.86	-0.5736
S	T4	0.0019	86.56	-0.4512
T	T1	0.0009	59.42	-0.2453
T	T3	0.0019	83.09	-0.3211
T	T4	0.0059	55.21	-0.7120

Participant ID	Time	α	β	ρ
U	T1	0.0016	60.28	-0.2433
U	T3	0.0041	11.61	-0.8500
U	T4	0.0008	93.56	-0.6122
V	T1	0.0015	57.70	-0.1091
V	T3	0.0014	77.039	-0.4418
V	T4	0.0008	92.83	-0.3187
W	T1	0.0007	50.94	-0.1423
W	T3	0.0013	125.5	-0.6085
W	T4	0.0212	1.336	-0.0419

Note. α = learning rate; β= inverse softmax temperature; ρ= perseveration.