MOTIVATIONAL CONTROL IN ATTENTION-DEFICIT/HYPERACTIVITY DISORDER:

A NEUROBEHAVIORAL AND TRANSLATIONAL ACCOUNT

by

AHMET OĞUZ CECELI

A Dissertation submitted to the

Graduate School—Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Psychology

written under the direction of

Dr. Elizabeth Tricomi

and approved by

_____

Elizabeth Tricomi

_____

Mauricio Delgado

_____

Miriam Rosenberg-Lee

_____

Catherine Myers

Newark, New Jersey

May, 2019

Copyright page:

ABSTRACT OF THE DISSERTATION

Motivational Control in Attention-Deficit/Hyperactivity Disorder:

A Neurobehavioral and Translational Account

By AHMET OĞUZ CECELI

Dissertation Director:
Elizabeth Tricomi

Abstract

Attention-deficit/hyperactivity disorder (ADHD) poses debilitating impairments in the neurobehavioral systems governing reward learning and processing—key components involved in the control of motivated behaviors. Specifically, ADHD may rely on a system favoring cue-driven habits—rooted in the posterior putamen—over caudate and prefrontal cortex-driven goal-directed behaviors. Impaired motivational control may accompany corticostriatal dysfunction in ADHD (e.g., altered connectivity and striatal recruitment). A comprehensive investigation of habits is necessary to reveal potential motivational control irregularities that may be associated with ADHD. However, although contemporary tools enable the study of habit formation, examining existing habits and their disruption has not garnered comparable interest, necessitating the development of novel methods to capture well-learned habits. This dissertation discusses

the neurobehavioral mechanisms of habit formation in ADHD in Study 1, develops a novel Go/NoGo task that capitalizes on existing green-Go and red-NoGo associations to study well-learned habit expression and disruption in Study 2, and applies these new tools to investigate habit expression and disruption as a function of ADHD symptomology in Study 3. In Study 1, despite similarities in behavioral assays of habit formation across groups, adults with ADHD displayed corticostriatal connectivity abnormalities and the hyper-recruitment of the posterior putamen during reward learning, alluding to a neural signature of impaired top-down control. In Study 2, participants exhibited outcome-insensitive habits when managing the Go/NoGo task, in that green-Go and red-NoGo associations elicited impairments in accuracy when incongruent with daily experiences. These habits were broken when participants were provided performance-tracking information paired with extrinsic reward, electing motivational enhancement via feedback as a candidate strategy for restoring goal-directed control. In Study 3, the novel task evoked well-learned habit expression and disruption independent of ADHD symptomology in the general population, although a modest association between hyperactivity and the prepotency to execute well-learned habits was evident. In sum, these studies suggest that ADHD presents corticostriatal abnormalities during motivational control, provide novel tools to better examine well-established habits and their disruption, and highlight the importance of investigating motivational systems in ADHD to better understanding its pathophysiology.

Acknowledgement

I am deeply indebted to several people who were instrumental in the completion of my doctoral training and dissertation.

First, I would like to express my most sincere gratitude to my advisor, Dr. Elizabeth “Eebie” Tricomi, for not only fueling my passion for science, but also for believing in me. Eebie’s mentorship propelled me to think critically about all aspects of my work, and instilled in me the much-needed confidence to overcome the highs and lows of graduate school. Her extensive knowledge and unwavering support made my graduate training an experience that I will always remember fondly.

I am also extremely grateful for the support of my dissertation committee. Dr. Mauricio Delgado played a critical role in my scientific training and professional development over the years. Mauricio has been a role model for me as a scientist way before I started this program, and I am tremendously grateful for his mentorship as my committee chair. My committee members Drs. Miriam Rosenberg-Lee and Catherine Myers improved not only the studies in this dissertation, but the way I approach research questions as a scientist. I feel incredibly lucky to have worked with these remarkable scholars.

In addition, I would like to express my appreciation for Dr. Luis Rivera’s many roles in my training. Luis was assigned to be my secondary mentor, helping me quickly adjust to graduate school when I began the program. Soon after, we became fellow committee members at the Diversity and Transformation Commission. Luis also served as a faculty advisor during my stint as the subject pool coordinator. In short, Luis has

provided invaluable support from my first to last days in Rutgers, for which I am especially grateful.

I would also like to thank all past and present members of the Learning and Decision Making Lab, along with other friends I have made over the years in the Psychology Department. I am grateful for having the support of Dr. Brynne DiMenichi, who was an exemplary labmate throughout my training. I thank Christina Bejjani, Dina Mattern, Christopher Cagna, and Chrystal Spencer for their invaluable support and insightful assistance in these studies. Special thanks go out to Dr. James Bradley, Leyla Roksan Caglar, Sally Wang, Dr. Catherine Cho, Dr. David Smith, and John Henes for their helpful advice and assistance in various stages of my dissertation work. I am fortunate to have developed such wonderful relationships in my time at Rutgers. I would like to also acknowledge my collaborators, Drs. Joman Natsheh, Daniel Cruz, and Jamil Bhanji for their valuable assistance, and the research assistants I have had the pleasure of working with over the years: Zana Hanini, Christine Oti, Sarah Ramirez, Charlie Ndouli, and Giavanna Esposito. The studies in this dissertation could not be completed without their help in recruitment and data collection.

My mentors from previous institutions certainly played a monumental role in my training. I would like to thank Dr. Richard Contrada for introducing me to research in psychology, Dr. Richard Ronay for his mentorship during my research assistant training in Columbia University, and Dr. Gerald Voelbel for introducing me to clinical neuroscience. I am especially grateful for having had the chance to work with Drs. Elizabeth Phelps and Joseph Dunsmoor at NYU. I attribute a great deal of my interest in studying cognitive neuroscience to Liz and Joey's mentorship.

I could not be even close to where I am without the undying love and support from my family and friends. This dissertation is dedicated to my parents, Tulin and Naci Ceceli, and my sisters Esra, Elif, and Evha. A special shout out goes to my friends for their constant encouragement, help with participant recruitment, and overall involvement in my life, especially throughout this journey.

Finally, I would like to thank all the participants who offered their time and appetite to partake in my studies. My research could not be possible without their generosity.

Table of Contents

List of Tables

List of Supplemental Tables

List of Illustrations

List of Supplemental Illustrations

Motivational Control in Attention-Deficit/Hyperactivity Disorder:
A Neurobehavioral and Translational Account

Chapter 1: Introduction

"*…the future does not exist for me—only the present…*"

When colloquially asked about how attention-deficit/hyperactivity disorder (ADHD) affected their daily functioning, my study participants disclosed scenarios as heterogeneous as the diagnostic criteria of ADHD. Several stories in line with the name of the disorder involved lost keys, ruined dates, missed birthdays, and impulsive decisions resulting in great emotional and financial distress. The handpicked quote above, however, struck me as peculiar—the participant had inadvertently summarized years of research that links ADHD to reward-related deficits. These studies will be examined in detail throughout this dissertation to situate a relevant yet understudied research avenue—the motivational underpinnings of ADHD. The quote serves as a reminder for the real-world implications of my work, and represents the overarching question I have attempted to answer during my doctoral training: do the reward-related deficits in ADHD extend to impairments in motivation, and if so, can function be restored? In other words, can the future (e.g., the motivation to pursue goals) become relevant in ADHD?

ADHD: a prevalent public health concern

ADHD is a prevalent childhood onset disorder afflicting 3.4% of individuals across the globe (Fayyad et al., 2007). The disorder is most notably associated with debilitating symptoms of inattention and hyperactivity (American Psychiatric Association, 2013). These symptoms manifest as fidgeting and restlessness, difficulty organizing and managing tasks, struggling to follow directions at work and/or school, and

a degree of distractibility that significantly hinders daily activities (Barkley, 2005). ADHD shows pronounced comorbidities with a wide range of psychiatric illnesses, such as oppositional defiant disorder, conduct disorder, substance use disorder, generalized anxiety disorder, autism spectrum disorder, and depression (Steinhausen et al., 2006). Fifty-two percent of individuals with ADHD are documented to display at least one comorbidity, while 26.2% are expected to possess two or more comorbidities (Jensen and Steinhausen, 2015). Not only is this prevalent childhood-onset disorder devastating for affected children, ADHD is known to persist into adulthood in approximately 38% of cases (Kessler et al., 2005). Therefore, studying adults with ADHD is paramount for understanding the disorder's mechanisms, and consequentially developing effective diagnostic and therapeutic strategies.

In the search for translational insight, examining the neural systems underlying ADHD is essential. Neurobiological investigations of ADHD have revealed reward-related abnormalities, illustrating that the disorder's scope of impact extends beyond attention and hyperactivity (Castellanos and Tannock, 2002). The experiments in this dissertation will test the reach of the reward-related deficits associated with ADHD by targeting the control of motivated behaviors. In three studies, I examine the neurobehavioral mechanisms underlying motivational control in ADHD, develop a novel tool to study and remediate impairments in motivational control, and deploy this method to better understand these processes in the context of ADHD symptomology.

Evidence of reward processing deficits in ADHD

In addition to the detrimental manifestations of attentional and impulsive symptoms impacting work, academic achievement, and interpersonal relationships,

ADHD has also been associated with reward learning impairments (Johansen et al., 2009). Several studies investigating cognitive function in children with ADHD have pinpointed difficulties adaptively processing reward (Douglas and Parry, 1983; Luman et al., 2008; Slusarek et al., 2001). Corroborating our participant's excerpt, developmental research has highlighted maladaptive delay discounting, such that children with ADHD exhibit delay aversion that results in choosing small immediate rewards at the expense of larger delayed rewards (Antrop et al., 2006; Kuntsi et al., 2001; Sonuga-Barke et al., 1992). Given the highly persistent nature of the disorder, these delay discounting abnormalities are also apparent in adults with ADHD (Kessler et al., 2005; Marx et al., 2010; Marx, Höpcke, Berger, Wandschneider, & Herpertz, 2013). Overall, ADHD has been characterized by a wide array of reward-related behavioral deficits, ranging from processing to interacting with rewarding outcomes.

ADHD and the brain's reward circuitry

The reward-related dysfunctions exhibited in ADHD are accompanied by abnormalities in the brain's reward circuitry (Castellanos and Tannock, 2002). In brief, the brain's reward circuitry regulates via cortico-striatal pathways the process of experiencing rewarding outcomes, learning from these rewards, and directing behaviors to maximize gain while minimizing loss (Daw et al., 2011; Delgado, 2007; Galvan et al., 2005; Knutson et al., 2001). Aberrant reward processing is largely reflected in the ADHD brain as decreased striatal signals during reward anticipation and altered orbitofrontal cortex (OFC) activation at reward receipt (Furukawa et al., 2014; Plichta et al., 2009; Plichta and Scheres, 2014; Scheres et al., 2007; Ströhle et al., 2008). Wilbertz and colleagues have shown that medial OFC activity tracks reward magnitude in

neurotypicals (NT), yet there exists a neural signature of overvaluing smaller rewards and undervaluing larger rewards in ADHD (Wilbertz et al., 2012). ADHD has been associated with pronounced OFC activity during reward delivery (Ströhle et al., 2008; von Rhein et al., 2015), though conflicting findings have also been reported, with no group differences in OFC recruitment (Stoy et al., 2011). Independent of directionality in OFC activation, the prevailing conclusion is that reward sensitivity in this region shows atypicality (Cubillo et al., 2012; Edel et al., 2013; Ströhle et al., 2008; von Rhein et al., 2015; Wilbertz et al., 2012), in cadence with the notion that ADHD is associated with a dysfunctional reward circuitry.

The evidence for irregularities in the brain's reward systems is further supported by studies examining neural connectivity and structural morphometry. ADHD is associated with heightened orbitofrontal-cingular communication (Tomasi and Volkow, 2012), as well as volumetric reductions in the anterior cingulate region (Carmona et al., 2005; Frodl and Skokauskas, 2012; Makris et al., 2007; Seidman et al., 2006, 2011). These cortical regions are imperative for adaptive reward processing and value-based decision making (Bush et al., 2002; Cole and Schneider, 2007). A basis for our motivational control stance on examining ADHD stems from such cortical anomalies. In short, ADHD's neural irregularities allude to a compromise in reward-driven interactions with the environment.

When discussing reward-related abnormalities, it is also worth considering ADHD's diagnostic profile, which comprises inattentive, hyperactive, and combined presentations (American Psychiatric Association, 2013). In particular, a topic of debate has been reward signaling differences across presentations. For instance, the hyperactive

symptom presentation has been associated with diminished ventral striatal activity during reward anticipation (Scheres et al., 2007). Yet, another report documented no such differences, with comparable neural signaling across ADHD and NT groups (Stoy et al., 2011). To further complicate the link between distinct presentations and neural signaling, researchers have later suggested that inattentiveness, not hyperactivity, is a key predictor of diminished ventral striatal signaling during reward anticipation (Edel et al., 2013). Extending presentation differences further, this study also reports hyper-responsiveness of the OFC during reward receipt in the combined presentation compared to the inattentive presentation. These inconsistencies in the neural signature of reward-related processes suggest that both inattentive and hyperactive domains may play a role in neurobehavioral systems in ADHD. Therefore, a symptom-based approach that can interrogate domain-specific contributions to potential deficits may be a valuable tool in examining this disorder.

Neural systems of motivational control: habitual and goal-directed behaviors

A critical region in the brain's reward circuitry—the striatum—is consistently implicated in the execution of motivated behaviors (Balleine and O'Doherty, 2009; Delgado, 2007; Dolan and Dayan, 2013; Knowlton and Patterson, 2016). In brief, the striatum receives dopaminergic input from the midbrain during appetitive events (O'Doherty et al., 2004; Schultz, 1997), and via the pallidum and thalamus, bidirectionally communicates with the prefrontal and motor cortices to drive actions that predict desirable outcomes (Haber, 2003) (see Figure 1). The dorsomedial portion of the striatum, known as the caudate in humans, forms connections with the prefrontal cortex to drive goal-directed behaviors that are performed in congruence with the value of a

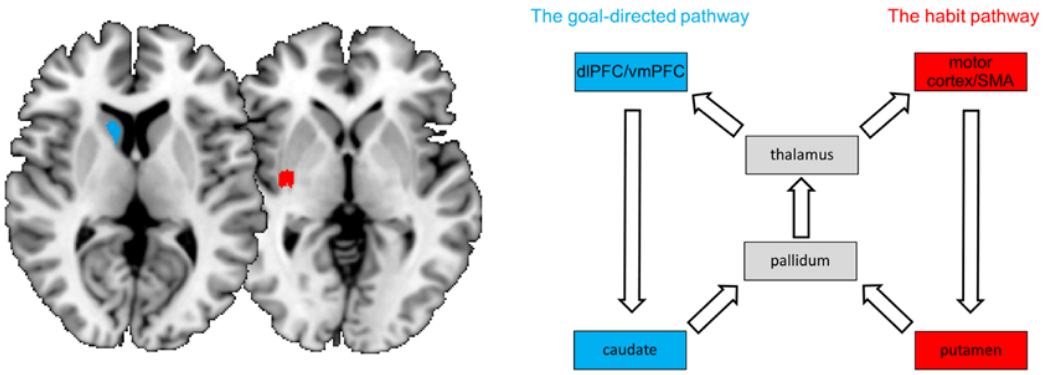


Figure 1. The corticostriatal pathways driving goal-directed and habitual control. Left: The posterior part of the putamen, highlighted in red, and the caudate nucleus, highlighted in blue. Right: A simplified schematic of the corticostriatal loops that regulate goal-directed and habitual control. Goal-directed actions rely on connectivity between the prefrontal cortex and the caudate, whereas habitual actions are regulated by a network of putamen and motor regions. dlPFC: dorsolateral prefrontal cortex; vmPFC: ventromedial prefrontal cortex; SMA: supplementary motor area. Figure and figure legend adapted from Ceceli & Tricomi, 2018.

consequential reward (Tricomi et al., 2004; Yin et al., 2005a, 2005b). The dorsolateral part of the striatum, referred to as the putamen in humans, connects with the motor cortex and the supplementary motor area, driving habitual behaviors that are triggered by salient, preceding cues, rather than the value of the contingent outcomes (Tricomi et al., 2009; Yin et al., 2004, 2006). The necessity of the prefrontal cortex and caudate for goal-directed behaviors, and the posterior putamen for the execution of habits has been consistently reinforced in the literature. Namely, ventromedial prefrontal cortex (vmPFC) lesions have been linked to motivational control that's dominated by outcome-insensitive habits (Reber et al., 2017). Multivariate pattern analyses of neural signals have pinpointed the vmPFC, dorsolateral PFC (dlPFC), and the caudate nucleus as regions that contrain response-outcome representations, while the posterior putamen has been shown to contain representations of stimulus-sensitive habits (McNamee et al., 2015).

Potential motivational control deficits in ADHD

Given the behavioral and neural reward processing deficits in ADHD, a fascinating yet elusive question remains: how is the motivational control of action that determines whether actions are habitual or goal-directed affected in this disorder? Indeed, the reward circuitry and motivational control literature rely on overlapping neural real estate, and ADHD is linked to functional and connectivity abnormalities in key corticostriatal regions (Costa Dias et al., 2013; Rosch et al., 2018; Tomasi and Volkow, 2012; von Rhein et al., 2017). Investigations of striatal and prefrontal morphometry provide further support for the notion that ADHD may be characterized by motivational control-related discrepancies. These studies reveal significant global gray matter volume and dlPFC reduction, caudate compression, and putamen expansion in children with ADHD relative to NTs. (Carmona et al., 2005; Qiu et al., 2009). Similarly, analyses of adult brains reveal structural discrepancies that may result in impaired goal-directed control, as diminished prefrontal cortical thickness is observed in ADHD (Makris et al., 2007).

Providing further support for the notion that ADHD may be characterized by deficits in motivational control, affected individuals also exhibit trait motivation deficiencies, which have been associated with midbrain and striatal dysfunction pertaining to dopamine receptor and transporter availability (Volkow et al., 2009, 2011). Recent work that more closely relates to the habitual and goal-directed components of motivational control has focused on a rat model of ADHD—the Spontaneously Hypertensive rat (SHR) strain that possesses ADHD-like symptoms of impulsivity (Natsheh and Shiflett, 2015). In this study, the SHR strain exhibits a hyperactive habitual

control system favoring cue-driven habits over outcome-driven, goal-directed behaviors. This finding has been extended to suggest that the reliance on habits in the SHR strain may be due to an imbalance in dopaminergic receptor activation (Natsheh and Shiflett, 2018), a conclusion that agrees with the finding that trait motivational impairments correlate with dopaminergic systems in ADHD (Volkow et al., 2011).

These functional, neurochemical, and structural disparities in brain regions that comprise the reward and motivational control circuitry suggest that ADHD may also be a disorder of motivational control. Furthermore, it can be argued that such neurobiological evidence for corticostriatal dysfunction foreshadows disparities involving stimulus-sensitivity, electing the posterior putamen and the prefrontal cortex key regions of interrogation. In Chapter 2, I posit that due to the strong neurobehavioral and neurochemical evidence for reward-related dysfunction, individuals with ADHD may be especially vulnerable to exhibiting maladaptive motivational control, potentially favoring putamen-driven habitual actions over deliberate, caudate and prefrontal cortex-driven, goal-directed execution.

Tackling motivational control in humans

To examine the neurobehavioral systems underlying motivational control in ADHD, we will need appropriate tools that successfully capture goal-directed and habitual processes. Habits are distinguished from goal-directed actions in that they are performed in response to a salient, triggering cue, without considering the outcome of this action (Dickinson and Balleine, 1994). Habits are thus ideal in instances where the agent needs to behave quickly (e.g., initiating the action of looking both ways before crossing the street, despite the absence of oncoming traffic). In contrast, goal-directed

behaviors require more cognitive resources, and are the product of value-based deliberation (e.g., choosing a restaurant for lunch) (Otto et al., 2013, 2015; Wood and Rünger, 2016). This intricate balance between the two components of motivational control, if compromised, may result in deleterious outcomes, such as sub-optimal decision making, and contribute to disorders of compulsion, such as obsessive-compulsive disorder and addiction (Everitt and Robbins, 2005; Gillan et al., 2016; Griffiths et al., 2014). Therefore, capturing a comprehensive snapshot of motivational control is paramount for better understanding the magnitude of potential neurobehavioral anomalies in ADHD. Accordingly, I approach the study of habits and goal-directed behaviors from three perspectives: (1) the formation of habits, (2) the expression of well-learned habits, and (3) the disruption of habits.

Examining habit formation—a brief overview of the contemporary methods

The motivational basis of habits and goal-directed actions have traditionally been studied using paradigms that introduce to subjects novel cue-action-outcome contingencies. One of the popular methods in detecting habits is the outcome-devaluation task where a primary or a secondary reward is devalued to test the behavior's outcome-sensitivity (Alvares et al., 2014, 2016; de Wit et al., 2012; Sjoerds et al., 2016; Tricomi et al., 2009). Another widely-used example is the sequential decision task, in which subjects respond to probabilistic multi-step associative sequences and recruit model-based (i.e., goal-directed; taking into account the cognitive model of the task environment) or model-free (i.e., purportedly habitual; selecting sub-optimal actions that are based solely on history of reward receipt) strategies to maximize gain and minimize loss (Daw et al., 2005). Additional tasks that capture different aspects of the habit experience have been

developed over the years, such as those that rely on stimulus pre-training for habit induction and contingency change for outcome-sensitivity testing (McKim et al., 2016). For a more detailed analysis of contemporary methods, please see (Ceceli and Tricomi, 2018; Watson and de Wit, 2018).

Studies that investigate the formation and expression of newly-developed habits are abundant in the literature, and the methods described above are all valid candidates for approaching habit formation in a clinical context. In Chapter 2, I tackle the first action step, the neurobehavioral mechanisms underlying motivational control in ADHD—specifically, the corticostriatal processes involved in habit formation. To this end, I deploy an outcome-devaluation task using food rewards, which allows us to monitor the neural changes throughout the potential development of stimulus-sensitivity in ADHD.

The gap in the literature: the expression of well-learned habits

Thanks to the methods described above, we have amassed a vast library of knowledge regarding habits. However, these paradigms remain insufficient in critical aspects. For example, tasks based on outcome-devaluation and sequential decision-making rely on the development of a newly-formed habit. This limitation renders our understanding of well-learned habits that are more representative of daily experiences disproportionately incomplete. In outcome devaluation paradigms, the cardinal test of habitual expression is an assessment of outcome-insensitivity. However, especially in studies involving food rewards, developing an outcome-insensitive habit relies on the over-training of a new cue-response-outcome association, subjective reports of selective satiety during the devaluation procedure, and assumptions of comparable food palatability (Tricomi et al., 2009). Not only is the experimental induction of habits

cumbersome, the resulting habits are weak associations that do not represent most real-world scenarios, and only provide a platform that captures the shift from goal-directed to habitual control (i.e., habit formation). Thus, we are limited in our tools to investigate well-learned habit expression and disruption. This is concerning for the habit scientist, given that the study of habit disruption can inform interventions applied towards the restoration of a compromised motivational control system (Ceceli and Tricomi, 2018).

To address this gap in the literature, in Chapter 3 I introduce our well-learned habit paradigm: a Go/NoGo task that capitalizes on the existing green-Go and red-NoGo associations that are assumed to be strengthened in a variety of contexts including traffic lights, signals of danger and safety, and childhood games, songs, and stories (Suskauer et al., 2008). In brief, this task assesses Go/NoGo performance when color-response mappings are congruent versus incongruent with daily experiences, exploiting the response bias that would potentially be triggered by familiar, congruent contingencies (i.e., green-Go, red-NoGo). Importantly, this approach permits the study of a well-established association that is strong enough to produce a meaningful investigation of habit disruption. Thus, several issues that plague the typical habit paradigm are circumvented: (1) ecological validity: the task capitalizes on existing associations that do not need extensive training; (2) objective assessment of outcome-sensitivity: the task does not rely on subjective reports of selective satiety—instead, outcome-sensitivity is tested by changing Go and NoGo contingencies, and calculating accuracy impairments that can be attributed to prepotent, habit-like responses; (3) habit strength: the task captures strong S–R associations that are impervious to changes in the environment,

facilitating the study of manipulations that can be used to restore goal-directed performance in potentially compromised populations.

Breaking the well-learned habit

Having situated the expression of well-learned habits in the framework of motivational control, a critical research question becomes apparent: how can these inflexible habits be disrupted? Because the driving force behind a habit is a salient stimulus dominating over outcome value, a candidate habit disruption strategy may involve amplifying the salience of the outcome. If the outcome representation can compete with the salience of the stimulus, the well-learned, stimulus-driven habits can be rendered flexible and goal-directed. In the context of my in-house Go/NoGo task, manipulations that allow the amplification of outcome value can be implemented into the task. If green-Go and red-NoGo associations are outcome-insensitive, in that performance is impaired when these contingencies are changed to reflect incongruent color-response associations, manipulations that prevent the incongruency-related impairments would mean an effective demonstration of habit disruption.

A motivational problem may call for a motivational solution. If habits are dominating over goal-directed behaviors due to a weaker representation of outcome value during action execution (O'Doherty, 2016), boosting motivation and focusing on enhancing outcome representations may prove useful in restoring goal-directed, value-driven control. Indeed, administering performance-based feedback (e.g., primary and secondary rewards) have been used extensively in enhancing behavioral performance (Kluger and DeNisi, 1996; Montague and Webber, 1965). Early investigations of performance-contingent feedback delivery (e.g., performance tracking information in the

form of a score or its combination with monetary incentives) successfully improved performance on a visual task (Montague and Webber, 1965). A combination of primary and secondary rewards (e.g., juice and monetary incentives) has improved goal-directed performance on a cued task-switching paradigm via motivational enhancement (Yee et al., 2016). The promise of a future reward contingent on performance has sufficed in improving performance during task-switching, and accelerating responses during a reaction time task with congruent and incongruent stimuli (Kleinsorge and Rinkenauer, 2012; Zedelius et al., 2012). The beneficial effects of feedback, especially transient monetary incentives (i.e., increasing reward magnitudes from low to high across trials) have survived against paradigms designed to tax executive control and visual perception (Shen and Chun, 2011). Due to the effects of performance-contingent feedback acting on the engagement of top-down control systems during task-switching (Umemoto and Holroyd, 2015), performance tracking and performance-contingent rewards may be prime candidates for enhancing goal-directed motivational control. In Chapter 3, I extend the well-learned habit expression focus to include the restoration of goal-directed performance via performance-contingent feedback (e.g., intrinsic and extrinsic rewards that promote behavioral flexibility).

Well-learned habit expression and disruption as a function of ADHD symptomology

One of the primary motivators for developing our well-learned habit task is the possibility of revealing goal-directed control impairments in potentially compromised populations, and taking a translational step in the direction of restoring function. To examine well-learned habit expression and disruption in the context of ADHD symptomology in Chapter 4, I apply the novel Go/NoGo task to a large sample of

participants from the general population, from whom I acquire ADHD-related symptom severity information. This approach enables us to examine whether ADHD symptom severity in the general population can track habitual control and the disruption of habits, paving the way for translational research avenues towards developing interventional strategies to promote behavioral flexibility. Another advantage of our symptom-based approach involves mapping our dependent variables to observable behavioral anomalies—the core philosophy behind the Research Domain Criteria (RDoC) project (Insel et al., 2010). I assess habit-related processes in the context of ADHD-related symptoms such as inattentiveness and hyperactivity, in absence of disorder-based exclusion criteria. Thus, I am able to improve the approach from Chapter 2 (and the typical study examining pathology via disorder classifications) by ensuring that the results are generalizable to a wider audience, as opposed to a carefully filtered subset of the population that experiences ADHD in absence of the prevalent comorbidities (McGough et al., 2005).

In sum, Chapter 2 tackles the neurobehavioral processes underlying ADHD and motivational control—specifically the formation of habits. Individuals with ADHD and matched NTs were scanned while they performed a reward-learning task and strengthened novel cue-action-outcome contingencies. This method investigated whether ADHD is associated with the proclivity to develop outcome-insensitive habits driven by altered cortical and striatal function in the brain. Next, in Chapter 3, I describe a task aimed to expand my scope of motivational control by employing more representative, well-learned habits that permit the study of habit disruption. To achieve this aim, I first developed and validated a novel Go/NoGo task that capitalizes on existing go and stop

habits elicited by green and red traffic lights. I then introduced a feedback-based manipulation that boosts motivation to disrupt the well-learned habits elicited by these stimuli. Lastly, in Chapter 4, I present a study on the process of well-learned habit expression and disruption as a function of ADHD symptom severity by applying the approach from Chapter 3 to a large sample of individuals from the general population.

Chapter 2: Investigating the neurobehavioral mechanisms underlying motivational control in ADHD

Introduction

Attention-deficit/hyperactivity disorder (ADHD) involves reward-related behavioral anomalies that significantly impair executive function and overall quality of life (Barkley, 1997; Castellanos and Tannock, 2002). ADHD is specifically associated with impaired reward learning (Johansen et al., 2009), difficulties adaptively processing rewards (Douglas and Parry, 1983; Luman Marjolein et al., 2008; Sethi et al., 2018; Slusarek et al., 2001), and heightened delay discounting (Antrop et al., 2006; Kessler et al., 2005a; Kuntsi et al., 2001; Marx et al., 2010, 2013; Sonuga-Barke et al., 1992). The cardinal symptom of impulsivity is a well-documented contributor to maladaptive reward-related behavioral rigidities such as addictions (Cunill et al., 2015; Urcelay and Dalley, 2012). We posit that these reward-related abnormalities may also drive aberrances in the neurobehavioral mechanisms underlying the control of motivated behaviors (i.e., striking an adaptive balance between cue-driven habits and outcome-driven, goal-directed behaviors).

A growing body of neurobiological evidence asserts that ADHD is also characterized by dysfunctions in the brain's reward circuitry (Castellanos and Tannock, 2002). This network of cortical (e.g., anterior cingulate cortex; ACC, ventromedial prefrontal cortex; vmPFC, orbitofrontal cortex; OFC) and sub-cortical (e.g., striatum, amygdala, and hippocampus) brain regions regulates the process of experiencing rewarding outcomes, learning from rewards, and directing behaviors to maximize gain while minimizing loss (Daw et al., 2011; Delgado, 2007; Galvan et al., 2005; Knutson et al., 2001). Compared to neurotypicals (NTs), individuals with ADHD exhibit

irregularities in reward-related neural processing, such as decreased striatal signals during the anticipation of a rewarding outcome, and increased orbitofrontal cortex activation at reward receipt (Furukawa et al., 2014; Plichta et al., 2009; Plichta and Scheres, 2014; Scheres et al., 2007; Ströhle et al., 2008).

The compromised neural systems that regulate reward-related processes largely overlap with the corticostriatal circuits that also drive motivated behaviors. Motivated behaviors are posited to be either controlled by the pursuit of a desirable outcome (i.e., deliberate and goal-directed), or triggered by an antecedent stimulus regardless of the outcome (i.e., reflexive and habitual; Adams, 1982; Dickinson and Balleine, 1994). These components of motivational control have distinct neural representations. The dorsomedial portion of the striatum (i.e., caudate in humans) forms connections with the prefrontal cortex to drive goal-directed behaviors that are performed in congruence with the value of a consequential reward (Tricomi et al., 2004; Yin et al., 2005a, 2005b). The dorsolateral part of the striatum (i.e., putamen in humans) fosters connectivity with the motor cortex and the supplementary motor area, guiding cue-based habits that are triggered by salient, preceding stimuli, rather than by the value of contingent outcomes (Tricomi et al., 2009; Yin et al., 2004, 2006). ADHD is associated with functional abnormalities in these key motivation-related regions. Attentional and motivational deficits in ADHD are correlated with disruptions in the dopaminergic reward pathways along the midbrain and striatum (Volkow et al., 2009, 2011). Importantly, ADHD is consistently associated with irregular fronto-striatal connectivity (Costa Dias et al., 2013; Rosch et al., 2018; Tomasi and Volkow, 2012; von Rhein et al., 2017). ADHD's potential motivational control abnormalities may present stimulus-sensitivity-related disparities,

making the posterior putamen and the prefrontal cortex candidates for aberrant signaling during associative learning and the strengthening of stimulus–response–outcome (S–R–O) associations. In support of this hypothesis, recent investigations of a rat model of ADHD have reported habit-driven action control in the Spontaneously Hypertensive rat (SHR) strain that possesses ADHD-like symptoms of impulsivity (Natsheh and Shiflett, 2015, 2018).

To examine the neural systems guiding habitual and goal-directed behaviors in ADHD, we administered a free-operant reward learning paradigm adapted from Tricomi et al. (2009) to individuals with ADHD and matched NTs who underwent functional MRI. We interrogated an *a priori* posterior putamen region of interest (ROI) to reveal potential differences in neural processing following moderate, single-day S–R training in ADHD. We also used the posterior putamen as a seed region in a psychophysiological interaction (PPI) to detect potential abnormalities in corticostriatal connectivity. Lastly, we employed a whole brain analysis to further examine the neural signature of motivational control in ADHD following moderate S–R training.

Materials and methods

Participants

A meta-analysis of studies investigating brain function in ADHD reports the average sample size per study as 28, including ADHD and neurotypical (NT) groups (Lei et al., 2015). Due to the heterogeneity in ADHD symptom severity and presentation (American Psychiatric Association, 2013; Ramtekkar et al., 2010), we increased our sample size to the upper range of the reviewed studies. Following the recruitment criteria outlined below, 25 adults with ADHD and 25 NTs matched on age, gender, handedness,

and working memory (WM) participated in the study (14 females and 11 males in each group; ADHD M_{age} = 22.32, SD_{age} = 4.69; NT M_{age} = 21.48, SD_{age} = 2.92, age range = 18-35; five left-handed participants in each group). WM was assessed via the Digit Span subtest of Wechsler's Adult Intelligence Scale (WAIS-IV; Wechsler, 2008). Independent samples t-tests were performed to ascertain that there were no significant group differences in age, $t(48) = 0.76$, $p = .451$, or WM, $t(48) = 0.40$, $p = .694$.

Informed consent was obtained from all participants per the ethical principles outlined on the Declaration of Helsinki (1964), and experimental protocols were approved by the Rutgers University Institutional Review Board.

Study Inclusion Criteria

Pre-interview screening

Individuals interested in participating in the study were provided pre-screener questionnaires via email to determine eligibility using Qualtrics (http://www.qualtrics.com). MRI-safe individuals (i.e., those without claustrophobia or ferrous metal in or on their bodies) were invited to undergo psychiatric interview only if they confirmed the absence of the following exclusion criteria: (1) neuropsychiatric illnesses or history of head injuries, (2) disqualifying psychoactive medication use (i.e., previous use of non-ADHD medication with psychoactive properties such as antidepressants and anxiolytics), (3) active dieting behaviors or concerns of body weight, and (4) reservations about consuming large amounts of chocolate and cheese crackers. Individuals who had received an ADHD diagnosis less than 12 months prior to the interview session were also restricted from participating in the study.

Psychiatric interview

All participants underwent a psychiatric interview session performed under the supervision of a clinician. Because most ADHD medications have a duration of action and half-life shorter than 12 hours (Kolar et al., 2008), individuals with ADHD were instructed to refrain from medication use for at least 24 hours prior to psychiatric assessment. During the psychiatric interview, measures assessing impulsivity via Barratt's Impulsiveness Scale (BIS; Patton *et al.*, 1995), ADHD symptom severity via ADHD Self-Report Scale (ASRS; Kessler et al., 2005), medication history, and working memory via WAIS-IV Digit Span were administered to detect individual differences for correlational analyses. The medication history variable served as an index of pharmacological treatment exposure and was scaled in months. Due to the snacks to be provided during the fMRI scan session, the Eating Attitudes Test 26 (EAT-26; Garner *et al.*, 1982) was administered to screen for maladaptive eating attitudes predictive of eating disorders. Participants with an EAT-26 score of 20 or above were excluded from the remainder of the study.

We followed the diagnostic criteria from Mini International Neuropsychiatric Interview 6.0 (MINI) Plus: Adult ADHD module (Sheehan et al., 1998) to confirm ADHD diagnoses, and the MINI 6.0 to rule out co-morbid psychiatric disorders (Sheehan et al., 1998). The psychiatric illnesses that served as exclusion criteria are as follows: major depressive disorder, manic-depressive disorder, generalized anxiety disorder, agoraphobia, social phobia, obsessive-compulsive disorder, post-traumatic stress disorder, alcohol and substance dependence/abuse, psychotic disorders, mood disorders with psychotic features, anorexia nervosa, bulimia nervosa, and antisocial personality

disorder. Participants were cleared for the MRI scan only if they were deemed free of psychiatric illness per MINI criteria, and above the clinical threshold of ADHD manifestation per the MINI Plus: ADHD module criteria.

MRI scan session

Following the psychiatric interview, qualifying participants were scheduled for the fMRI scan. Participants were instructed to fast for at least 4 hours prior to the scan session to increase the desirability of the snacks to be used as rewards throughout the study. Participants with ADHD were instructed to refrain from taking ADHD medication for at least 36 hours before the scan to prevent the acute effects of these medications from affecting the results.

Experimental paradigm

E-prime (Psychology Software Tools, Pittsburgh, PA) was used for stimulus presentation and response collection. Prior to entering the MRI scanner, we collected subjective pleasantness ratings for each snack to be used during the free-operant task. Specifically, because the task involved rewarding actions with M&M (Mars, McLean, VA) and Goldfish cracker (Pepperidge Farm, Norwalk, CT) outcomes, participants were asked how pleasant they would find eating an M&M and a Goldfish cracker on a scale of 0 to 5.

Next, after completing a brief practice session outside of the scanner, participants underwent a free-operant task with food rewards during fMRI, similar to Tricomi et al. (2009). In this paradigm, two “task” fractals predicted differential snack outcomes contingent on button press responses, such that index and middle finger button presses produced either M&M or Goldfish outcomes. A third fractal was used to indicate

unrewarded rest trials, for which participants refrained from making any response. Specifically, participants were informed that fractal images would be presented throughout the experiment, and a schematic above each fractal would indicate which button was activated for response collection for that fractal. Participants were instructed that during each trial in which a fractal was presented with an active button, they could respond via button presses as often as desired to earn the associated snacks, and that they should pay attention to the fractal–response–snack associations. Each active button response produced either a gray circle (50 ms) or a snack image corresponding to the snack earned (1 s) below the fractal. Rewards were administered on a random interval reinforcement schedule (RI-10), meaning each second, a participant had a 0.1 probability of earning a reward following a button press. Thus, a reward became available on average every 10 seconds, and was collected upon the first response executed by the participant following its availability (see Figure 2 for task structure). This RI reinforcement schedule has been shown to be conducive to developing outcome-insensitivity when compared to fixed or variable-ratio reward delivery (Baum, 1993; Knowlton and Patterson, 2016). The fractal–button and button–snack associations were counterbalanced across subjects. The responses towards task fractals were self-paced, and fractal onset and offset indicated the start and end of each trial. Twelve task (six of each fractal, randomly varying durations of 20, 30, or 40 seconds) and eight rest trials (20 seconds) comprised each 8-minute run.

Participants underwent a moderate amount of training (six runs) and were then taken out of the MRI scanner for the outcome-devaluation procedure.

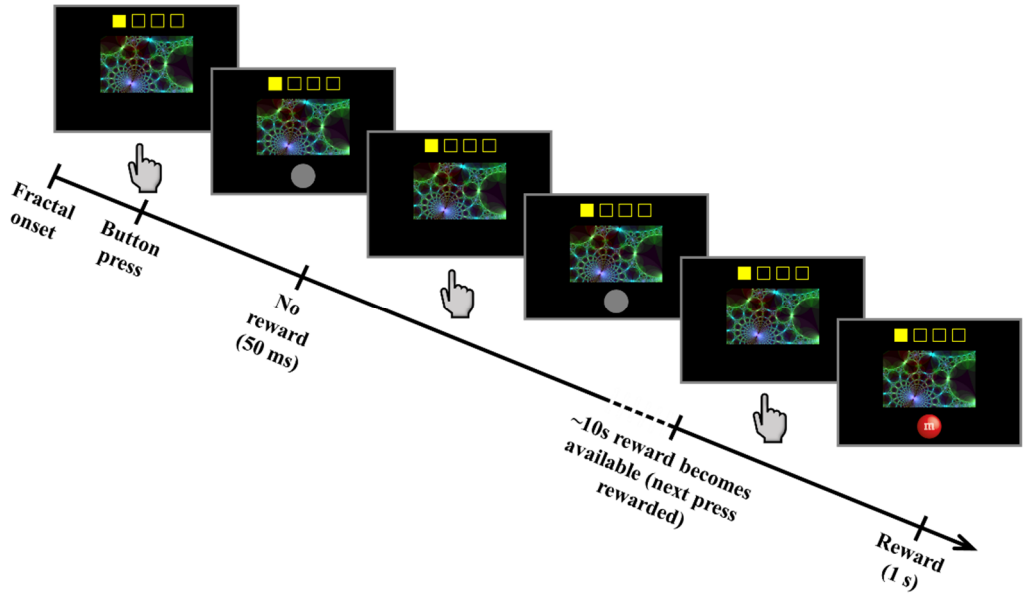


Figure 2. Free-operant task structure. Participants were trained under a RI-10 reinforcement schedule using two distinct S–R–O associations. Fractal stimuli were paired with a yellow schematic that served as an indicator of the active button. A reward was made available on average every 10 seconds. Participants were trained for six runs while undergoing fMRI, then taken out of the scanner to complete a devaluation procedure. Following the devaluation of one of two snacks, we placed participants in the scanner again, and tested in extinction whether they persistently responded to the stimulus that predicted the now-devalued reward, indicating outcome-insensitivity.

Snack earnings accumulated throughout the task were given to participants at a ratio of 4 images to 1 snack to prevent satiety. One of the snack outcomes was then made available to the participant until it was no longer pleasant, effectively diminishing its value. Specifically, the participant was instructed to consume the snack until it was no longer pleasant. The experimenter remained in the room with the participant during the devaluation procedure. The snack chosen for selective satiety was counterbalanced across subjects. Post-devaluation subjective ratings of snack pleasantness were collected to ensure that the now-devalued snack outcome was perceived as less valuable compared to the pre-training ratings. The experimenter offered more of the snack to the participant if

these ratings did not decrease from their pre-training responses. Following successful devaluation, participants re-entered the fMRI context and underwent an identical free-operant task, but unbeknownst to them, the trials were no longer rewarded. This three-minute extinction stage allowed us to determine whether button presses were outcome-sensitive. For instance, diminished button press responses to the fractal associated with the now-devalued snack would indicate outcome-driven behavior, as the participant's response rate slows down in accordance with the outcome value. In contrast, a persistent response rate to the stimulus predictive of the devalued snack would indicate stimulus-driven performance, as the participant responds at a similar rate regardless of snack value. The extinction test was unrewarded to prevent newly acquired rewards from affecting the outcome-sensitivity measures, and took place during fMRI to provide similar training and testing contexts.

FMRI data acquisition

A 3 Tesla Siemens Trio (Erlangen, Germany) MRI scanner with a 12-channel phased array coil was used to acquire structural and functional brain images at the Rutgers University Brain Imaging Center (RUBIC). High-resolution T1-weighted structural images at isotropic 1mm voxel dimensions were obtained using a Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence. Forty-one T2*-weighted echo-planar image slices were obtained for blood oxygenation level dependent (BOLD) signal analyses using the following parameters in an interleaved order of acquisition: 3 mm isotropic voxels, TR: 2500 ms TE: 30 ms, field of view: 192 mm x 192 mm, flip angle: 90°). We acquired brain images at a 30° oblique orientation to the anterior commissure – posterior commissure axis to improve signal-to-noise ratio,

particularly in the ventral prefrontal cortex region that is most susceptible to signal dropout (Deichmann et al., 2003).

Data analysis

Behavioral data analysis

To determine whether ADHD and NT groups exhibited differential motivational control in response to outcome devaluation, we performed a mixed-design ANOVA with the dependent variable (DV) as the change in response rate (ΔResponse_Rate) between training and extinction (extinction minus training responses per second), Stimulus Value (valued, devalued) as a within-subjects, and Group (ADHD, NT) as a between-subjects factor. Post-hoc t-tests were used to confirm whether ΔResponse_Rate to valued and devalued stimuli were significantly different in each group. Pleasantness rating comparisons between pre-training and post-extinction stages were performed using paired-samples t-tests to ensure that the outcome-devaluation procedure was successful in diminishing the subjective value of the devalued snack.

Although we matched ADHD and NT groups on a variety of individual difference measures (i.e., demographics and WM capacity), we also aimed to further examine our sample's diagnostic profile. We performed a multiple regression analysis with Devalued_ΔResponse_Rate as an index of behavioral flexibility. We used symptom severity scores obtained via the ASRS survey, impulsivity scores via BIS, WM via Digit Span, and treatment history as indexed by years of medication use as regressors to detect relationships with devaluation-related changes in response rate to the devalued cue (Devalued_ΔResponse_Rate) in the ADHD group. A similar model using only the Symptom, Impulsivity, and WM variables as regressors were used with NT data to

determine whether sub-clinical ADHD symptom severity, impulsivity, and WM predicts behavioral flexibility in the NT sample. This analysis served as an exploration of individual variability within our sample, and these variables' potential links to motivational control.

FMRI data analysis

We used FSL (version 5.0; http://www.fmrib.ox.ac.uk/fsl) for fMRI data pre-processing and analysis. We skull-stripped brain images to eliminate non-brain matter from analyses and employed FMRIB's Linear Image Registration Tool (FLIRT) to spatially transform our functional and structural images to the Montreal Neurological Institute (MNI) template (Jenkinson & Smith, 2001). We accounted for subject movement via FSL's MCFLIRT (Jenkinson & Smith, 2001; Jenkinson et al., 2002) and extracted six motion parameters to be included as regressors of no interest in the general linear model (GLM). Additionally, we identified volumes that showed spikes in translation and rotation parameters when compared to a reference volume. These outliers were determined by a typical boxplot threshold (75th percentile + 1.5 * interquartile range) using fsl_motion_outliers and regressed out in the GLM. This algorithm removed an average of 5.8% of the volumes in each run (range: 1%-16.7%). Following this outlier removal procedure, no substantial volume-to-volume movement remained, in that maximum displacement was below the voxel dimensions (mean motion: 0.33 mm; maximum motion: 1.26). Importantly, neither the number of spikes in movement, $t(295) = -0.52$, $p = .600$, nor the volume to volume displacement, $t(48) = -0.64$, $p = 525$, was significantly different across ADHD and NT groups. The BOLD data we acquired in an interleaved order were slice-time corrected and spatially smoothed using a 5 mm full

width at half maximum Gaussian smoothing kernel, and a high-pass filter cutoff of 100 s was applied to ignore extraneous signal from the imaging context. Three runs (ADHD = 2, NT = 1) were discarded due to data corruption, and the GLM was altered accordingly for these participants' analyses. These three runs were in the middle of their respective scan sessions, such that the analyses aimed at deriving late and early S–R activation estimates were unaffected.

We generated parameter estimates from the pre-processed data for each subject using a GLM approach. "Task" onset and "rest" onset were captured via 1-second events at each trial, and a "reward" regressor was captured as a 1-second event at reward receipt. We did not include extinction scan data in the GLM, as these brief scans do not provide sufficient power for fMRI data analysis. Therefore, all analyses of neural data inform processes involved in moderate S–R learning, but not devaluation or extinction. These task, rest, and reward regressors, their temporal derivatives, along with the six motion parameters and motion outlier timeseries were convolved with a canonical hemodynamic response function (HRF). Linear contrasts of task versus rest onset were computed in each run to selectively determine stimulus-evoked activation patterns. Each subject's first-level parameter estimates were entered into a fixed-effects model to generate subject-level "early," "mid," and "late" regressors (two runs in each bin), denoting the stage of S–R learning during the free-operant task. These learning phase regressors were parametrically weighed (-1, 0, 1 as early, mid, and late) and used in the group-level region of interest (ROI), psychophysiological interaction (PPI), and whole-brain analyses outlined below.

ROI analysis: posterior putamen and stimulus-sensitivity

Due to the strong *a priori* hypothesis centered on the role of the posterior putamen in driving stimulus-sensitivity, we created 5 mm-radius spherical anatomical masks of left and right putamen (±33, -24, 0) using MNI coordinates obtained from a previous study employing the same free-operant task (Tricomi et al., 2009). We extracted percent signal change values from this posterior putamen seed region in each subject's early, mid, and late stage, task versus rest contrast image using FSL's Featquery tool. We performed a repeated measures ANOVA using Time (early, mid, late) as a within-subjects, and Group (ADHD, NT) as a between-subjects factor to detect posterior putamen activation differences across groups as a function of training length (i.e., a Group x Time interaction). Given the incremental recruitment of the posterior putamen over the course of extended S–R training in the general population (Tricomi et al., 2009), a similar pattern over moderate training in the ADHD group would suggest an accelerated recruitment of this region closely associated with stimulus-sensitivity.

To further examine our sample's diagnostic profile in the context of stimulus-sensitivity-related neural signaling, we also performed a multiple regression analysis to reveal potential associations between treatment history, ADHD symptom severity, impulsivity, and working memory (WM) capacity on posterior-putamen activity in each group. The regressors Medication, Symptom, Impulsivity, and WM were used to predict percent signal change in the posterior putamen ROI over the course of training in the ADHD group. The regression was repeated for the NT group without the inclusion of the Medication variable. This set of analyses served as an exploration of individual

variability within our sample, and these variables' potential links to motivational control-related striatal signaling.

Psychophysiological interaction: posterior putamen functional connectivity

In a PPI analysis, we used our *a priori* posterior putamen region as a seed to identify target brain areas that exhibited functional connectivity as a result of moderate S–R training (Friston et al., 1997). This approach permitted us to assert whether corticostriatal discrepancies across ADHD and NT groups exist over the course of moderate S–R learning. We concatenated our task and rest regressors into a single time course, weighed task as 1 and rest as -1 to create a "psychological" regressor which was convolved with an HRF. We also included this regressor's temporal derivative in the GLM. Next, we extracted timeseries information from our anatomically extracted 5-mm left posterior putamen mask, which was transformed into each subject's functional space, to create a "physiological" regressor. Using these two regressors, we derived a "PPI" (i.e., interaction of psychological and physiological time courses) regressor in our design matrix, Lastly, we also included in our PPI GLM regressors representing all other events throughout the fMRI scan: an HRF-convolved regressor that weighed the task and rest events equally (i.e., as they would be represented in the whole-brain GLM), an HRF-convolved reward regressor for reward receipt events, and motion outlier parameters. This method allowed us to estimate target brain regions that exhibited task-based coupling with the posterior putamen seed, while ruling out regions that may evoke continuous, non-task-specific coupling, such as anatomical connectivity (O'Reilly et al., 2012).

The second-level analysis, as outlined in the FMRI Data Analysis section above, aggregated the first two runs to derive “early”, the middle two runs to derive “mid”, and the last two runs to derive “late” run regressors at the subject level. We assigned parametric weights to these regressors (-1, 0, 1), and performed linear contrasts of the resulting statistical maps to calculate late versus early activation patterns in each participant. For the group-level estimation of moderate S–R learning-related functional connectivity (ADHD versus NT group contrast), we entered these maps into a mixed-effects model using FLAME 1 & 2 (FMRIB's Local Analysis of Mixed Effects), which performs Markov Chain Monte Carlo simulations to improve variance estimation and permit population inferences (Beckmann et al., 2003). We employed a cluster defining threshold of $p < .005$, corrected to a cluster extent threshold of $p < .05$ (greater than 207 contiguous voxels to constitute a significant cluster).

Whole-brain GLM

To identify brain regions involved in moderate S–R training in ADHD, we derived parametrically weighed early, mid, and late training phase activation maps from each participant and entered them into a mixed-effects model using FLAME 1 & 2, with a cluster defining threshold of $p < .005$, corrected to a cluster extent threshold of $p < .05$ (greater than 302 contiguous voxels to constitute a significant cluster). These stimulus-driven statistical maps denoted which brain regions were significantly active over the course of moderate S–R learning, and allowed for ADHD versus NT contrasts for group-level comparisons to elucidate potential differences in the ADHD brain.

Results

Sample profile

Table 1. Study sample profile.

	ADHD	NT	sig.
Age (SD)	22.32 (4.69)	21.48 (2.92)	p = .451
Sex	F=11, M=14	F = 11, M=14	individually matched
Handedness	R=20, L=5	R=20, L=5	individually matched
Working Memory: Digit Span (SD)	28.16 (3.86)	27.60 (5.93)	p = .694
Impulsivity: BIS (SD)	75.92 (5.89)	69.40 (6.19)	< 0.001
Symptom Severity: ASRS (SD)	45.76 (9.01)	14.24 (7.51)	< 0.001

The ADHD group scored significantly higher in measures of impulsivity, $t(48) = -3.81$, $p < .001$, and symptom severity, $t(48) = -13.44$, $p < .001$ (see Table 1 for details). Eighteen adults with ADHD were currently medicated, 4 were previously medicated, and 3 were medication-naïve. Eight adults with ADHD received some form of psychological therapy from a clinician; 5 of these individuals underwent cognitive-behavioral therapy. The NT group reported no history of medication use or psychological therapy.

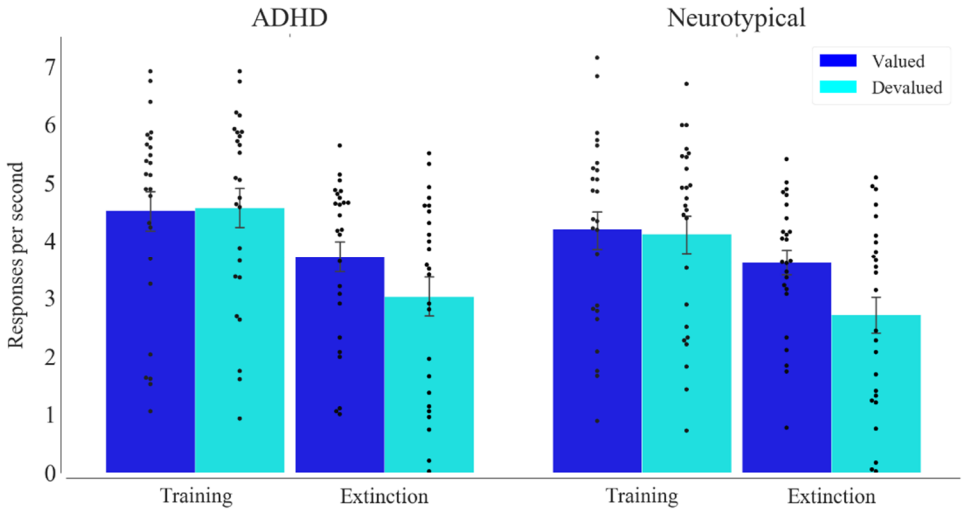


Figure 3. Response rate pre-devaluation (Training) and post-devaluation (Extinction). ADHD and NT groups exhibit similar sensitivity to outcome value. Behavioral similarities here suggest that both groups maintained goal-directed control (i.e., diminished responses to the devalued stimulus at extinction). Error bars depict standard error of the mean. Swarm plot points represent data from individual subjects.

Behavioral results

We did not find a significant difference in outcome-sensitivity across ADHD and NT groups. The mixed-design ANOVA of ΔResponse_Rate revealed no main effect of Group, $F(1,48) = 0.24$, $p = .629$, $\eta_p^2 < .01$, a main effect of Stimulus Value, $F(1,48) = 10.76$, $p = .002$, $\eta_p^2 = .06$, and no Group * Stimulus Value interaction, $F(1,48) = 0.02$, $p = .876$, $\eta_p^2 < .01$, suggesting that the ADHD and NT groups did not differ in devaluation sensitivity and both performed in a goal-directed manner (see Figure 3). Pre- versus post-devaluation comparison of pleasantness ratings confirmed that across all participants, the devaluation procedure successfully diminished the subjective value of the devalued snack, $t(46) = 16.00$, $p < .001$.

We performed a multiple regression analysis in our ADHD group data, using the variables Symptom, Impulsivity, Working Memory and Medication to predict our subject-level measure of outcome sensitivity, Devalued_ΔResponse_Rate. None of these regressors significantly predicted outcome-sensitivity as assessed by Devalued_ΔResponse_Rate (all p-values > .291). Similarly, when this regression analysis was repeated in our NT group data to detect individual differences in sub-clinical ADHD symptom severity, impulsivity, and working memory, no regressor significantly predicted outcome-sensitivity (all p-values > .161). These results suggest that individual variability in our sample's diagnostic profile did not significantly contribute to behavioral flexibility.

FMRI results

ROI analysis: posterior putamen and stimulus-sensitivity

The posterior putamen has been shown to play a role in stimulus-sensitivity in the general population over extended (i.e., 3-day) S–R training (Tricomi et al., 2009). We

tested whether ADHD is associated with an accelerated recruitment of the posterior putamen as a result of single-day moderate training via a mixed-design ANOVA. We extracted BOLD data from a posterior putamen mask (Figure 4A) and used the percent signal change as DV, Group as a between-subjects, and Time as a within-subjects factor. In the left posterior putamen, we found no main effect of Group, $F(1,48) = 0.20$, $p = .657$, $\eta_p^2 < .01$, no main effect of Time, $F(2,96) = 0.73$, $p = .482$, $\eta_p^2 = .01$, but a significant Group * Time interaction, $F(2,96) = 5.35$, $p = .006$, $\eta_p^2 = .10$ (Figure 4B), suggesting that the ADHD group exhibits heightened posterior putamen activity as a function of training length. In the right posterior putamen, we did not find a significant main effect of Group, $F(1,48) = 0.25$, $p = .618$, $\eta_p^2 < .01$, no main effect of Time, $F(2, 96) = 0.1\ 1$, $p = .897$,

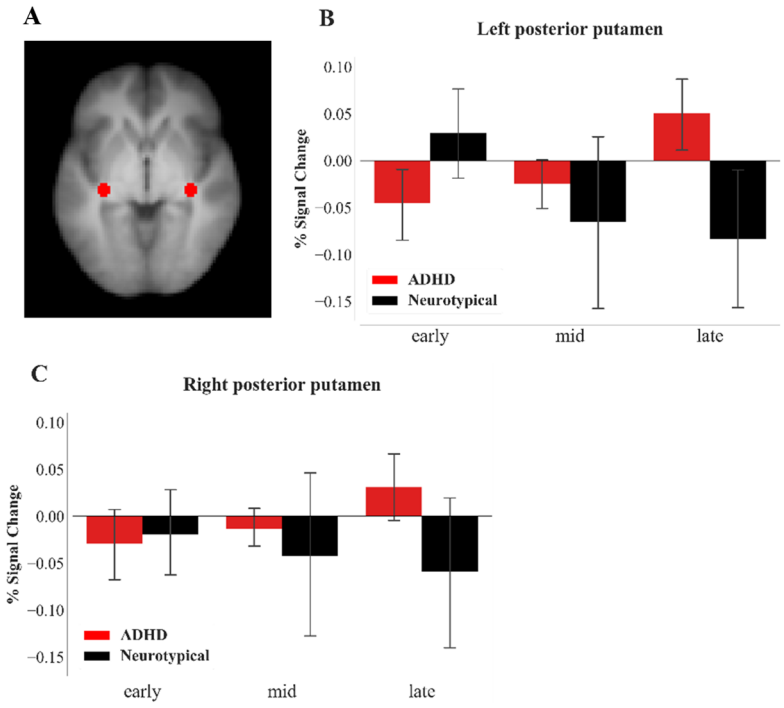


Figure 4. Posterior putamen ROI activity as a function of training length. A: We extracted left and right posterior putamen ROIs using coordinates from a previous study that associated this region with tracking stimulus-sensitivity. Percent change in BOLD signal is depicted from both hemispheres at MNI coordinates: ±33, -24, 0. B: Left posterior putamen significantly increases in activity as a function of S–R training length in ADHD ($p = .006$). C: No significant Group * Time interaction observed in the right posterior putamen ROI ($p = .257$).

$\eta_p^2 < .01$, and no significant Group * Time interaction, $F(2,96) = 1.38$, $p = .257$, $\eta_p^2 = .03$ (Figure 4C). The heightened recruitment of the left posterior putamen in the ADHD group may be due to ADHD being associated with an early onset of stimulus-sensitivity-related neural signaling, whereas this process may come online with more S–R training in the general population.

We performed a multiple regression using the late versus early contrast percent signal change derived from the parametrically weighed S–R training parameters (-1 0 1 as early, mid, late). We used this left posterior putamen ROI signal change as DV, and individual difference measures of medication history, symptom severity, impulsivity, and WM as regressors. In the ADHD group, we found no associations between any of these regressors and the percent signal change extracted from our left posterior putamen ROI (all p-values > .452). Similarly in our NT group, we found no associations between these regressors and left posterior putamen activity following moderate S–R training (all p-values > .207). These results suggest that posterior putamen activation over the course of S–R training was not affected by individual variability in our sample's diagnostic profile.

Psychophysiological interaction: posterior putamen functional connectivity

Considering the reward circuitry irregularities in ADHD (Castellanos and Tannock, 2002), we hypothesized that corticostriatal communication may be impaired during associative learning in ADHD. We performed a PPI analysis using our posterior putamen ROI—a striatal sub-region that has been regarded to play a key role in stimulus-sensitivity—as a seed and searched for areas that exhibited significant task-related coupling. We found that over the course of moderate S–R training, the posterior putamen fostered diminishing functional connectivity with the prefrontal cortex in the ADHD group when compared to the NT; namely in the dorsal anterior cingulate cortex (dACC)

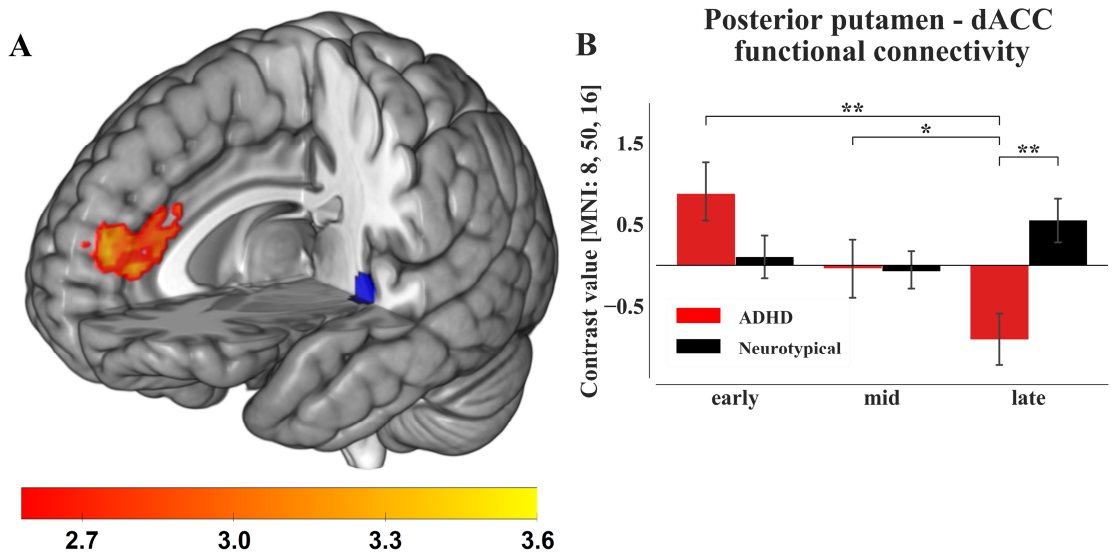


Figure 5. PPI analysis reveals corticostriatal connectivity differences in ADHD. A: We found diminished task-onset-related functional connectivity between the posterior putamen seed (in blue)—a striatal region that's been associated with tracking stimulus-sensitivity—and the dACC/mPFC (in warm colors) over the course of S–R learning. These prefrontal areas are known to be involved in error detection, reward value tracking, and goal-directed control. B: Peak voxels from this PPI analysis show decreased strength in corticostriatal connectivity in the ADHD group compared to NTs. Contrast depicted: task versus rest onset, late versus early phase; early, mid, and late training phases weighed as parametric regressors [-1 0 1]. Contrast values are derived from a 3 mm mask of peak activation (MNI coordinates: 8, 50, 16).

and in the medial prefrontal cortex (mPFC; see Figure 5. For the specific coordinates of clusters and local maxima associated with the PPI, see Appendix, Supplemental Table 1.

Whole-brain GLM

We performed a whole-brain analysis using the GLM approach to identify regions that drive stimulus-sensitivity following moderate S–R learning in ADHD. Specifically, we calculated a linear contrast of task versus rest onset to extract stimulus-sensitivity-related activity, then parametrically weighed the early, mid, and late stages of training (as -1 0 1, respectively) to examine stimulus-sensitivity over the course of moderate S–R learning. Lastly, we performed a group-level contrast of ADHD versus NT group to distinguish ADHD-specific stimulus-sensitivity-related brain activity following learning (see Figure 6). We found significant clusters in the posterior putamen, opercular/insular

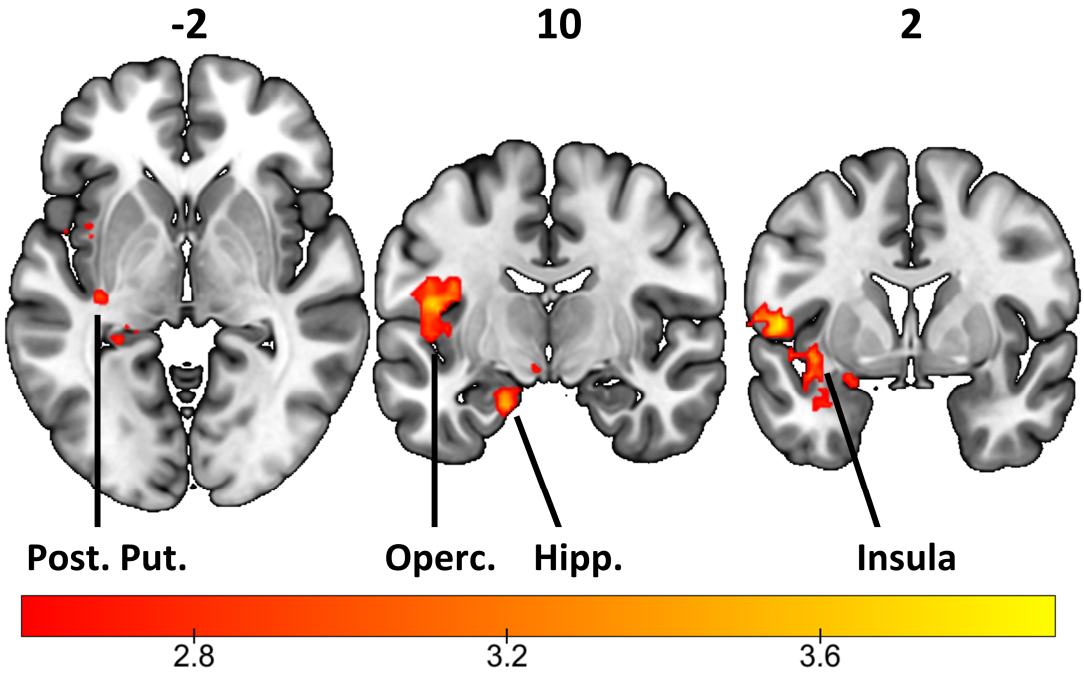


Figure 6. Whole brain analysis of cue-sensitivity over the course of moderate S–R learning in ADHD. Significant clusters in the posterior putamen, opercular/insular cortex, and the hippocampus survived our group-level comparisons of training-related cue-sensitivity activity. Clusters were defined using a Z-threshold > 2.58 ($p < .005$), corrected to the cluster extent threshold of $p < .05$. Numbers above brain slices indicate MNI coordinates. Post. Put: posterior putamen, Operc: operculum, Hipp: hippocampus.

cortex, and the hippocampus that selectively activated in the ADHD group. The posterior putamen region that survived our thresholding parameters overlaps with our *a priori* ROI,

which was based on Tricomi *et al.* (2009). For specific coordinates of clusters and local maxima associated with the whole-brain analysis, see Appendix, Supplemental Table 1.

Discussion

We investigated the neural signature of motivational control in ADHD by moderately training S–R associations during fMRI. This approach allowed us to identify corticostriatal abnormalities associated with the stimulus- and outcome-driven control of action in ADHD, interrogating the neural mechanisms underlying motivation-related processes. Namely, despite the intact goal-directed control following moderate S–R training in ADHD, we found an early recruitment of a striatal sub-region that has been associated with the execution of stimulus-sensitive behaviors (i.e., posterior putamen). We also revealed the posterior putamen's deficient connectivity with the prefrontal cortex in our ADHD sample. Additionally, our findings allude to a hippocampal process that could potentially be compensating for the corticostriatal abnormalities in order to maintain goal-directed control.

In the general population, the posterior putamen has been reported to show increased activation following over-training in a similar reward-learning paradigm (Tricomi et al., 2009). Here, we show that despite behavioral similarities across groups indicating outcome-sensitive, goal-directed control, an early recruitment of this posterior putamen region is evident in ADHD. The neural signature found in the ADHD group following a single-day's exposure to the S–R–O associations is similar to the over-trained participants in the report by Tricomi and colleagues (2009).

The prefrontal cortex—namely the dACC and vmPFC—are regarded as playing major roles in goal-directed control. These fronto-cingular areas have been identified as

components of the brain's cognitive control network (Cole and Schneider, 2007; MacDonald et al., 2000). An aberrance in this system may relinquish control in motivated behaviors to render them automatic and habitual (Otto et al., 2015; Poldrack et al., 2005; Verbruggen and Logan, 2009). Our PPI results reveal a diminished functional connectivity between the posterior putamen sub-region of the striatum and the dACC/mPFC. Given the dACC's role in cognitive control (MacDonald et al., 2000), error detection (Garavan et al., 2002; Polli et al., 2005), and reward-based decision making (Bush et al., 2002), its altered connectivity with the striatum following moderate S–R training may be indicative of sub-optimal neural processing that underlies value-driven action execution in ADHD. Along with the vmPFC—a prefrontal sub-region that is associated with value-tracking and inhibition/reversal learning (Smith et al., 2010; Zhang et al., 2016), the dACC's abnormal connectivity with the posterior putamen in ADHD may be an important biomarker for potential aberrances in value-based decision making and goal-directed control.

ADHD has been previously linked to ACC dysfunctions. Dampened inhibitory control-related ACC activity has been reported in adults with ADHD (Schneider et al., 2010). Structural abnormalities have also been documented, as ADHD is associated with volumetric reductions in the anterior cingulate region (Carmona et al., 2005; Frodl and Skokauskas, 2012; Makris et al., 2007; Seidman et al., 2006, 2011). Healthy individuals foster an antiphasic connectivity between task-positive (e.g., the cognitive control network including the dACC) and task-negative (e.g., the default mode network—a set of brain regions that coactivate in task-negative contexts) areas (Cole et al., 2014). However, adults with ADHD have been documented to display an impairment in this

functional connectivity between cognitive control and default mode regions, possibly driving the attentional lapses associated with the disorder (Castellanos et al., 2008). Children with ADHD, on the other hand, display heightened fronto-cingular connectivity patterns, in that the orbitofrontal cortex—a region associated with salience attribution and reward representations (Schultz et al., 2000; Sescousse et al., 2010)—fosters increased connectivity with the dACC (Tomasi and Volkow, 2012). This reward-related irregular connectivity may manifest as motivational deficits in ADHD, such as unfavorable value-based decision making. Our finding of a diminished corticostriatal connectivity—specifically, the communication between the posterior putamen and the dACC, may similarly allude to motivational impairments. The dACC is an important node for attentional processes such as response monitoring and selection (Bush et al., 1999; Camille et al., 2011). A compromised connection between regions driving stimulus-sensitivity, such as the posterior putamen, and response selection, such as the dACC, may produce action execution that is biased towards salient, triggering stimuli rather than towards outcome value.

Interestingly, we also see evidence for a hyper-recruitment of the hippocampus at task-onset following moderate S–R training in ADHD. The hippocampus is regarded as integral for declarative learning, and especially critical for contextual memory (Chun and Phelps, 1999; Greene et al., 2007). Possibly, the fractals in each trial may also provide contextual information, in that the indicators above the fractal that signal the active button may take on the properties of a stimulus, and the fractal may serve as a context in which the stimulus signals a response-contingent reward. The ADHD-specific hippocampal activation may therefore relate to the contextual information provided by

the fractals, which may be aiding in the maintenance of an outcome-driven behavioral profile despite the corticostriatal abnormalities. An interesting avenue for future research may be to further dissociate the potential hippocampal compensatory mechanisms that underlie aberrant prefrontal and striatal control systems.

Our whole brain analysis yielded stimulus-related activations in the opercular/insular region following moderate S–R learning. The insular recruitment in the late stage of S–R learning may be related to its role in the maintenance of rigid behaviors. The insula plays a critical role in addiction maintenance, in that damage to the insula predicts addiction disruption (Naqvi et al., 2007; Naqvi and Bechara, 2010). Thus, the insula may be an important player in developing rigid actions that eventually become outcome-insensitive habits. Future investigations of the insula and associative learning in ADHD can further elucidate the necessity of this region in executing stimulus-dependent actions. Furthermore, particular sub-regions of the insular and opercular cortex have been implicated in gustatory processes—even those involving the imagined taste of a stimulus (Barrós-Loscertales et al., 2012; Veldhuizen et al., 2007). The pronounced opercular activity in ADHD may be evoked by the enhanced encoding of the fractal cue as a contextual marker that predicts food rewards. Indeed, the coordinates that were associated with these processes in previous research overlap with our opercular and insular activation patterns (Barrós-Loscertales et al., 2012; Veldhuizen et al., 2007). Possibly, the food reward-associated fractal cues evoke gustatory processes more strongly in the ADHD group compared to NT.

Our devaluation procedure resulted in comparable sensitivity to the value of the outcome in both groups, suggesting the maintenance of goal-directed control across the

board. Although moderately training novel S–R–O associations was effective in identifying atypical brain function during reward learning, these behavioral similarities suggest that a direct investigation of habitual control in ADHD may require prolonged S–R training, or tasks that capture well-learned habits that do not rely on the traditional measures of devaluation sensitivity (Ceceli et al., submitted). Nonetheless, the neural findings reported in this study do not depend on devaluation or the extinction test, as we focused on the neural signature of moderate S–R learning during the training phase. Any heterogeneity in participants' behavioral sensitivity to devaluation is independent from the late stage training—the period of interest for neural calculations of associative learning strength. The corticostriatal abnormalities reported here are evident when the S–R–O associations in both groups should have moderately strengthened in late training.

ADHD is a highly prevalent disorder, and the wide range of reward-related irregularities warrants a closer examination of neurobehavioral mechanisms. We contribute to the growing neurobiological evidence for reward- and motivation-related dysfunctions in ADHD by highlighting key corticostriatal abnormalities affecting the posterior putamen, dACC, and mPFC during motivational control. Importantly, the atypical neural signaling related to motivational processes may indicate an ADHD endophenotype. Research on habits and goals in ADHD is imperative to better elucidate the neurobehavioral systems of habitual control, and ultimately advance our understanding of ADHD's neural anomalies to identify biomarkers in this debilitating disorder.

Chapter 3: Demonstrating and disrupting the execution of well-learned habits (submitted)

Introduction

When categorizing motivated behaviors, habits are distinguished from goal-directed actions in that they are performed reflexively in response to a triggering cue, without consideration of the consequences (Dickinson and Balleine, 1994). These habitual behaviors are less cognitively taxing than their goal-directed counterparts, allowing for their utilization in instances where the resource-consuming reflection of potential outcomes may not be ideal (Otto et al., 2013, 2015; Wood and Rünger, 2016). For example, looking both ways before crossing a street is an action best elicited habitually, and ideally should persist despite the absence of oncoming traffic. In contrast, the optimal motivational control system for commuting to a new destination would be outcome-reliant, reflective, and thus resource-consuming goal-directed performance.

For decades, the motivational bases of behavioral control (i.e., goal-directed and habitual actions) have been investigated in rodent models. In a typical study examining habitual control, a neutral stimulus (e.g., a visual cue, or the context of the chamber) signals hungry rats to press a lever in pursuit of a food outcome. This behavioral training period is often followed by a devaluation procedure—the rat is allowed free-access to the food, promoting satiation and diminishing the food's value (hence the term *devaluation*). In a subsequent, unrewarded, extinction phase, the experimenter can then assess whether the trained lever-press action is flexible and goal-directed (i.e., strong responses when animal is hungry but diminished responses when satiated), or rigid and habitual (i.e., persistent responses regardless of satiation) (Adams and Dickinson, 1981). Generally, over-training of the stimulus–response–outcome association tends to render actions

habitual. Thus, an over-trained rat persists in pressing the lever despite a diminished value in outcome, suggesting that the actions are driven by the preceding cue or the chamber context. In contrast, value-driven goal-directed control survives following moderate experience with the stimulus–response–outcome chain (Adams, 1982). Motivational control testing in humans has followed suit with similar operant conditioning paradigms, in which a primary or a secondary reward is devalued to determine whether actions are cue or value driven (Alvares et al., 2014, 2016; de Wit et al., 2012; Sjoerds et al., 2016; Tricomi et al., 2009; Valentin et al., 2007). Another widely-used example is the sequential decision task, in which subjects respond to probabilistic multi-step associative sequences and recruit model-based (i.e., goal-directed; taking into account the cognitive model of the task environment) or model-free (i.e., similar to habits; actions based solely on history of reward receipt) strategies to maximize gain and minimize loss (Daw et al., 2005).

These methods have undoubtedly contributed a great deal to our understanding of habits; however, such paradigms are limited in critical aspects. First, in contemporary paradigms, including those based on outcome-devaluation and sequential decision-making, the agent must develop a newly formed habit. Accordingly, the tools at our disposal facilitate the study of novel, lab-developed habits, while leaving incomplete our understanding of well-learned habits that are more representative of daily experiences. For example, especially in outcome-devaluation tasks involving valued and devalued food rewards, testing whether a behavior is habitual relies on several critical factors. The demonstration of a habit may depend on successful over-training of a new cue–response–outcome association that develops a strong enough link between the cue and the response

to guide behavior (Tricomi et al., 2009). Furthermore, the effectiveness of the devaluation procedure where a food outcome is selectively fed to diminish its value may become problematic in humans for reasons not encountered in rats, such as demand characteristics, and hesitation to eat copious amounts of junk food in a potentially socially intimidating lab setting. Lastly, the experimenter makes assumptions of comparable food palatability, in that the subject must value the food options similarly prior to selective devaluation for any value-based manipulation to be effective (Tricomi et al., 2009). These lab-generated habits are also arduous to develop via over-training, especially in expensive neuroimaging contexts. More importantly, the strength of the trained habit would be insufficient for a meaningful investigation of the habit-breaking process, in that even multi-day training is often measured in minutes to hours (McKim et al., 2016; Tricomi et al., 2009). Thus, the current tools provide a costly platform that only captures the unidirectional shift from goal-directed to habitual control (Ceceli and Tricomi, 2018). In other words, although these novel, lab-created associations permit the study of habit formation and execution, we are limited in our tools to investigate habit disruption with similar efficacy.

Despite tremendous efforts directed towards understanding habit formation and expression, a wider gap in the literature remains regarding the breaking of habits. Accessing the shift from habitual to goal-directed control may ultimately facilitate interventions that remediate rigid and maladaptive behaviors, yet we are not currently methodologically equipped to tackle this translational research avenue with a rich toolkit. Accordingly, we propose that developing a novel habit from an action–outcome contingency is not a pre-requisite for studying the motivational basis for habits, but that

an existing, more robust habit could be examined in the lab with less effort. An effective approach may involve using salient cues that elicit well-established, habit-like behaviors that are impervious to their consequences. For instance, the colors red and green have highly specific “stop” and “go” associations, possibly strengthened in a variety of contexts including traffic lights, visual signals of danger and safety, and childhood games, songs, and stories (Suskauer et al., 2008). The familiar red–stop and green–go contingencies have previously been transformed into Go/NoGo tasks to assess response inhibition via perseverative errors (i.e., NoGo accuracy) (Mostofsky et al., 2003; Naito and Matsumura, 1996; Suskauer et al., 2008). Similarly, we can test for behavioral rigidity by assessing performance when these contingencies are congruent with daily experiences versus when adjusted to reflect outcomes incongruent with most real-world scenarios. Thus, instead of devaluing the palatability of a primary reward, we render a well-learned association inappropriate for optimal task performance. The agent must override a prepotent red stimulus–stop response with an incongruent green stimulus–stop response to achieve the intended, correct outcome. A more pronounced accuracy impairment when managing incongruencies within this well-learned color-response mapping, compared to changes in a newly-acquired mapping, would permit us to conclude that these familiar stimuli evoke outcome-insensitive actions, the hallmark of habitual behavior. Upon establishing that these familiar stimuli elicit habitual control, we can then provide the platform to study habit disruption by testing manipulations that protect against mapping-related performance impairments–essentially preventing habitual control. The motivational control framework identifies habits as cue-dependent, and goal-directed behaviors as those contingent on the outcome (Dickinson and Balleine, 1994).

Accordingly, a previously goal-directed behavior is rendered habitual when the associative strength of the stimulus–response component governs actions, rendering the outcome inessential for action execution. A promising strategy for restoring goal-directed control may be via boosting the salience of the outcome—for instance, by enhancing the link between the response and outcome.

Providing opportunities for performance tracking and administering other forms of performance-based feedback (e.g., primary and secondary rewards) have been used extensively in enhancing behavioral output (Kluger and DeNisi, 1996; Montague and Webber, 1965). For instance, the delivery of performance tracking information combined with a monetary reward successfully improved performance on a visual task (Montague and Webber, 1965). A combination of primary and secondary rewards (e.g., juice and monetary incentives) has also been documented to improve goal-directed performance on a cued task-switching paradigm via motivational enhancement (Yee et al., 2016). The promise of a future reward contingent on performance has sufficed in improving performance during task-switching, and accelerating responses during a reaction time task with congruent and incongruent stimuli (Kleinsorge and Rinkenauer, 2012; Zedelius et al., 2012). Furthermore, trial-by-trial, transient monetary incentives (i.e., increasing reward magnitudes from low to high across trials) have served as salient performance boosters in tasks that taxed executive control, as well as visual perception (Shen and Chun, 2011). Taken together with the finding that performance-contingent monetary rewards engage top-down control on task-switching (Umemoto and Holroyd, 2015), performance tracking and performance-contingent rewards may be prime candidates for enhancing goal-directed behavioral control. Thus, we propose that boosting motivation

via performance-contingent feedback (e.g., intrinsic and extrinsic rewards that promote task performance improvements) may serve as a useful tool in restoring flexibility in otherwise rigid behaviors.

To achieve the goal of demonstrating and breaking a well-established habit, we introduce in Experiment 1 our novel Go/NoGo task that capitalizes on the familiar Green–Go, Red–NoGo associations people typically develop throughout the course of their lives. If the red–stop and green–go associations are well-learned, outcome-insensitive habits, there should be within-subject decrements in performance on an incongruent mapping of color to response (green–stop, red–go) compared to the well-learned congruent mapping (red–stop, green–go). In comparison, there should be no such within-subject differences between conditions involving novel color–response mappings (e.g. blue–stop, purple–go vs. purple–stop, blue–go). That is, if participants are responding habitually, they should be more likely to make errors of commission (i.e., responding to a cue when instructed to withhold responding), than if they are responding in a goal-directed manner. Then, in Experiments 2 and 3, we explore strategies to disrupt these well-learned habits by amplifying the salience of the action outcomes. Specifically, we use cumulative performance-contingent feedback to disrupt the incongruency-related impairment—which would restore goal-directed control in the face of habit-eliciting stimuli by reducing outcome-insensitive responses.

Experiment 1

Methods

Participants

We recruited 50 undergraduate students (32 female, 18 male participants; M_{age}=20.28, SD_{age}=2.96) from the Rutgers University-Newark campus for course credit. All subjects provided informed consent. Study protocols were approved by the Rutgers University Institutional Review Board. Participants were excluded if they reported having color-blindness.

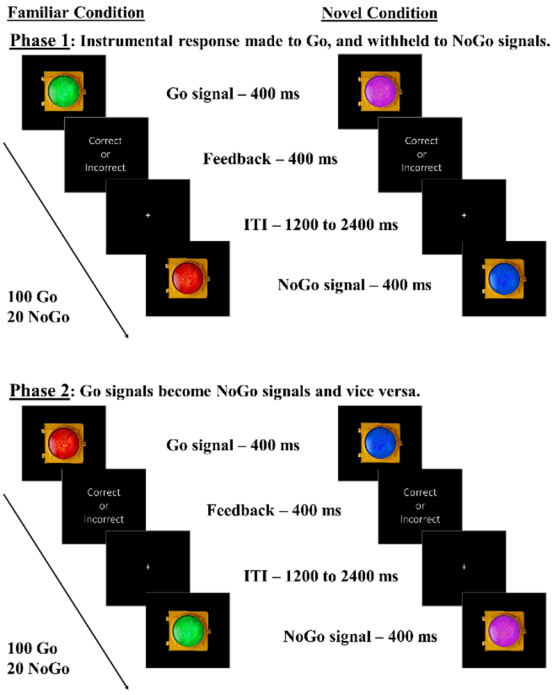


Figure 7. Go/NoGo task with familiar and novel lights. Participants are assigned to Familiar or Novel conditions. In the Familiar condition, subjects complete two phases: one where green signals Go and red signals NoGo ("congruent" mapping) and one where red signals Go and green signals NoGo ("incongruent" mapping). In the Novel condition, participants complete two similar phases, but the colors are blue and purple, for which there should be no strong pre-existing associations with "stop" and "go" responses. We predicted more commission errors in the Familiar condition for incongruent than congruent mappings, indicating outcome insensitivity, with no such within-subject differences expected in the Novel condition. Phase orders were counterbalanced across subjects.

Materials and Procedures

Participants were administered the Barratt Impulsivity Scale (BIS) (Patton et al., 1995), and randomly assigned to one of two stimulus type conditions (Familiar or Novel stimuli). They underwent a Go/NoGo task in which either Green and Red (Familiar condition) or Purple and Blue (Novel condition) traffic lights comprised Go and NoGo signals. Participants were instructed to respond as quickly and accurately to these stimuli as possible using the keyboard. A second phase followed in which the color-response

contingencies were swapped (see Figure 7). Note that in the Familiar condition, the Green-Go/Red–NoGo mapping was considered "congruent" with associations in everyday life, while the Red–Go/Green–NoGo mapping was considered "incongruent." We assumed that the Novel stimuli have no well-established Go or NoGo associations in daily life. The order in which participants underwent the two phases of the task was counterbalanced to ensure that the results could not be attributed to a specific order of managing the contingencies. Thus, we were able to examine the rigidity of our Familiar behavioral contingencies hypothesized to elicit outcome-insensitive responses in relation to a Novel stimulus set. An exit survey with demographic information concluded the study (see Figure 8 for a schematic of the experimental design).

Each phase comprised 100 Go and 20 NoGo trials (5:1 Go-NoGo ratio). The Go/NoGo stimuli remained onscreen for 400 milliseconds (ms), and each response produced a brief "correct" or "incorrect" text slide that offset after 400 ms. Go responses

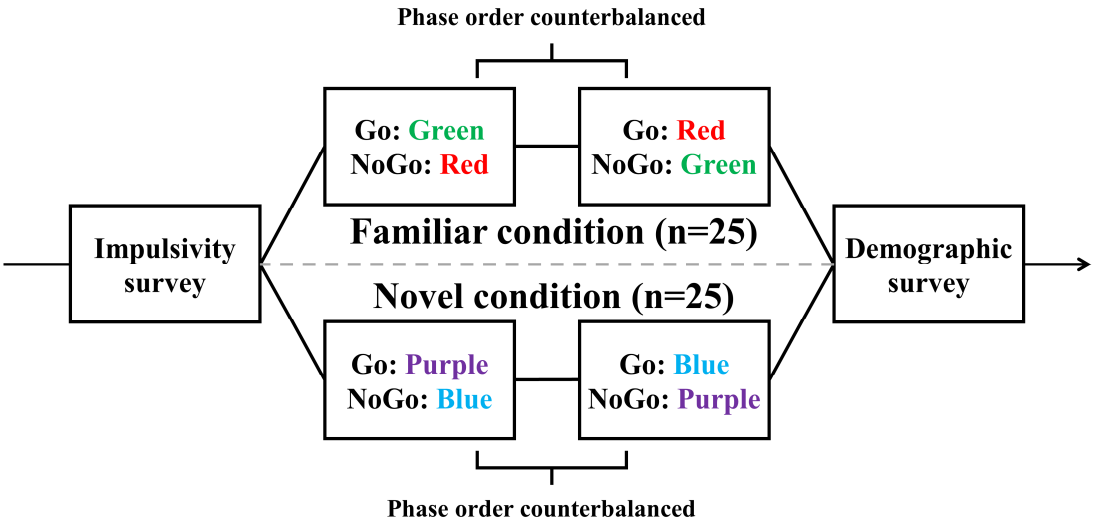


Figure 8. Experiment 1 design. Participants complete an impulsivity survey, then undergo either the Familiar or the Novel conditions of the Go/NoGo task. Participants perform both color-response mapping phases, and the order in which these contingencies are managed is counterbalanced across participants. A demographic survey concludes the experiment.

had to be performed before stimulus offset to be registered as correct. The inter-trial intervals varied randomly between 1200 and 2400 ms to ensure engagement with the task. All participants completed a brief practice session (six correct Go or NoGo responses) using the same stimuli as the first phase. This practice session was conducted with the experimenter present to ensure the comprehension of instructions.

If these familiar associations elicit habitual, cue-driven behavioral control, subjects undergoing the Go/NoGo task in the Familiar condition should experience a significant impairment in NoGo accuracy when incongruent with lifelong experiences (Green–NoGo). Accordingly, because the Novel condition stimuli are not characteristic of strong Go or NoGo signals, participants should show similar performance levels for both color–response mappings, illustrating the flexibility of responses executed towards the novel stimuli.

Data Analysis

Because the high ratio of Go to NoGo signals was expected to produce pre-potent Go responses, NoGo accuracy served as the primary measure of interest. As a secondary measure of outcome-sensitivity, identical analyses were performed using Go accuracy as dependent variable (DV). A mixed ANOVA with a DV of NoGo accuracy, Condition (Familiar or Novel stimulus conditions) as a between-subjects factor, and Mapping (congruent or incongruent mapping in the Familiar, and arbitrary color–response mapping in the Novel condition) as a within-subjects factor, was performed using Age, Gender, and Impulsivity (BIS score) as covariates. Post-hoc t-tests were employed to detect mapping-related differences in both conditions. We also performed a confirmatory omnibus test containing information from both conditions—a hierarchical multiple

regression to test the predictive strength of the Condition variable on mapping-related impairment. We summarize these omnibus regression data below, but refer readers to the Appendix for details (Supplemental Tables 2 and 3).

To determine sample size for our study, we performed an *a priori* power analysis using the effect size from an existing study examining Go/NoGo contingency change (Finn et al., 1999). A within-group comparison of commission errors due to contingency change—one similar to the primary analyses reported above—determined that 12 participants would be needed per group to reach 80% statistical power. We adjusted this sample size in accordance with our two between-subjects factors that yielded four groups, (two Condition levels and two Order levels – that is, the counterbalanced orders in which participants completed the two phases of the task), warranting a sample size of 50.

Results

Primary index of outcome-sensitivity: NoGo accuracy

To examine whether condition (Familiar or Novel) predicted outcome-sensitivity, we performed a repeated measures ANOVA using NoGo accuracy as DV, Condition as a between-subjects factor, Mapping as a within-subjects factor, controlling for Age, Gender, and Impulsivity as covariates. We found no main effect of Condition, $F(1,45) = 0.99$, $p = .325$, $\eta_p^2 = .02$, or Mapping, $F(1,45) = 0.10$, $p = .748$, $\eta_p^2 < .01$. but as evident in Figure 9, we found a significant Condition x Mapping interaction. $F(1,45) = 8.65$, $p = .005$, $\eta_p^2 = .16$. Post-hoc paired-samples t-tests further revealed a significant difference in NoGo accuracy in the Familiar condition, $t(24) = 3.53$, $p = .002$, suggesting that the incongruent "Green–NoGo" association elicits more errors of commission, indicative of outcome-insensitive, habitual control. Contingency change yielded no differences in

errors of commission between phases in the Novel condition, supporting the labile nature of newly learned associations, $t(24) = -0.88$, $p = .387$.

The omnibus regression test confirmed the significant effect of Condition. When controlling for participants' age, gender, and self-reported impulsivity, the inclusion of the Condition regressor in the hierarchical multiple regression model explained an additional 15.5% of the variance in outcome-sensitivity: $\beta_{Condition} = -0.40$, $p = .006$, $\Delta R^2 = .15$, indicating differential outcome-sensitivity across Familiar and Novel conditions. The details of this omnibus regression test and beta weights of all model parameters can be found in the Appendix (Supplemental Table 2).

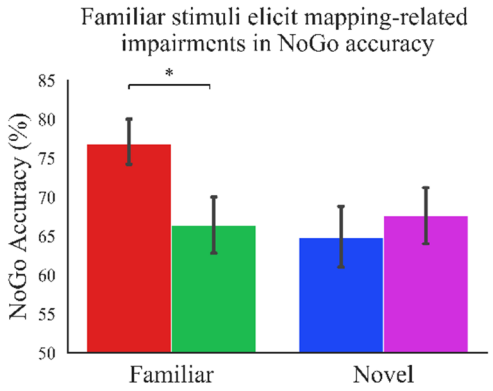


Figure 9. Familiar stimuli elicit mapping-related impairments in NoGo accuracy. Subjects make significantly more errors of commission when the NoGo signal is red compared to green. There is no difference in accuracy in the Novel condition when the NoGo signal is purple vs. blue. Condition x Mapping interaction: $p = .005$. Error bars depict standard error of mean (SEM). Color of bars reflects NoGo stimulus colors.

Secondary index of outcome-sensitivity: Go accuracy

A mixed-design ANOVA controlling for Age, Gender, and Impulsivity as covariates, using Go accuracy as DV revealed no main effect of Condition, $F(1,45) = 0.19$, $p = .667$, $\eta_p^2 < .01$, or Mapping, $F(1,45) = 2.93$, $p = 0.094$, $\eta_p^2 = .06$, but a Condition x Mapping interaction at $F(1,44) = 3.93$, $p = .054$, $\eta_p^2 = .08$ (Figure 10). Post-hoc paired-samples t-tests suggested a Go accuracy impairment only in the Familiar condition, $t(24) = 3.10$, $p = .005$, whereas no mapping-related Go impairment was observed in the Novel condition, $t(24) = 0.28$, $p = .785$.

The omnibus regression test also confirmed the role played by Condition on our secondary assay of outcome-sensitivity, Go accuracy. Controlling for participants' age, gender, and impulsivity scores, the inclusion of the Condition regressor significantly predicted mapping-related Go accuracy changes: $\beta_{Condition}$ = -0.27, ΔR^2 = .07, p = .049 (see Supplemental Table 3 in the Appendix for details).

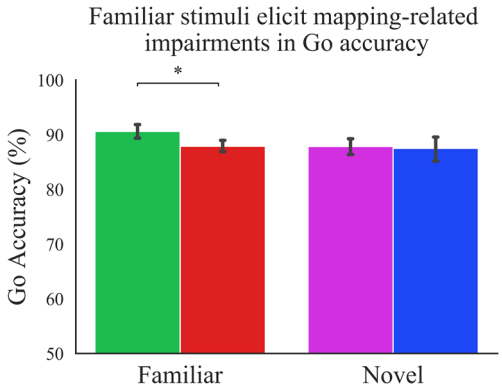


Figure 10. Familiar stimuli elicit mapping-related impairments in Go accuracy. Subjects perform worse when the Go signal is red compared to green. No such differences are seen in the Novel condition, when Go signal is blue vs. purple. Condition x Mapping interaction: p = .054. Error bars depict standard error of mean. Color of bars reflects Go stimulus colors.

These Go accuracy data lend support to the hypothesis that while red and green stimuli are rigid and habitual in triggering stop/go actions, blue and purple stimuli are not strongly associated with behavioral significance, in that they are labile and sensitive to the changes in action-outcome contingencies.

Discussion

This experiment demonstrates that habitual behavior that capitalizes on existing, non-lab-derived associations, can be demonstrated in the lab. By using the strong links between the green–go and red–stop associations in a Go/NoGo task, we were able to quantify the degree of flexibility to well-stamped in cue–response–outcome associations. Importantly, our results suggest that responses are more outcome-insensitive (i.e., habitual) when the stimulus meanings are congruent with our experiences with traffic lights in daily life (i.e., when a traffic light indicating "stop" is red, rather than green, blue or purple). We note that incongruency-related impairments alone are not enough to

conclude that a response is habitual; rather this conclusion must be verified by a comparison of the habitual associations (i.e., green–go, red–stop) with the novel control condition Go/NoGo associations (i.e., purple–go, blue–stop). Specifically, these red and green light stimuli triggered outcome-insensitive actions as evidenced by an accuracy impairment when Go and NoGo contingencies were incongruent with their well-established meanings outside of the lab. In contrast, the novel purple–go and blue–stop contingencies are not well-established in one's daily experiences, and their associative strength is limited to the participant's brief experience in the lab. Therefore, compared to the familiar stimuli, the actions evoked by the novel stimuli are more flexible to contingency changes, as reflected by similar NoGo and Go accuracy scores for blue vs. purple.

Assessing motivational control, which attributes the source of one's actions to either a preceding cue or its consequences, has long relied on experimental manipulations of outcome value. Rodent and human studies employing outcome-devaluation procedures of food rewards have depended on the subjects' comparable palatability of the foods used in the research, as well as the development of an outcome-insensitive habit via over-training of these action-outcome contingencies (Dickinson et al., 1995; Tricomi et al., 2009). Other researchers have made use of the instructed devaluation of outcomes, and computational investigations of choice strategy categorizations of model-based and model-free performance (Sjoerds et al., 2016). Although tremendously effective in their own avenues, a common area outside of the reach of these tasks is well-learned habits that better represent real world scenarios.

Our Go/NoGo task with familiar and novel stimuli provides new possibilities in studying habits. We demonstrate habits in a lab setting using stimuli that do not require lengthy training sessions to develop strong stimulus–response associations. This time- and cost-effective paradigm can serve as an especially useful tool in studying habits in expensive neuroimaging contexts. Perhaps more importantly, taking advantage of well stamped-in cue–response associations to study habits promises to contribute to translational science via new research avenues. For instance, although contemporary paradigms have proved fruitful in studying the formation and expression of habits, the nature of the tasks do not facilitate the investigation of habit disruption. Novel associations that have become outcome-insensitive following limited, lab-specific experience may not be rigid enough to represent real-world behaviors, and breaking these weak habits may not be translationally valuable.

We attempt the breaking of well-learned habits in Experiment 2, in which we boost motivation via cumulative performance feedback prior to contingency reversal. Because the motivational control framework attributes habits to be driven by antecedent cues and goal-directed actions to be guided by resulting outcomes, we hypothesized that amplifying the salience of the outcome may promote goal-directed performance at the expense of habitual control, thus aiding in breaking the well-learned habit.

Experiment 2

Methods

Participants

We recruited 100 undergraduate students (67 female and 33 male participants; M_{Age}=20.26, SD_{Age}=3.05) from the Rutgers University-Newark campus. All participants

provided informed consent and received course credit for their participation. Study protocols were approved by the Rutgers University Institutional Review Board. Participants were excluded if they reported having color-blindness.

Procedures

For the Go/NoGo task, participants were randomly assigned to a Feedback Group or No Feedback Group, and within each group, participants were randomly assigned to either Novel or Familiar condition, as in Experiment 1.

Feedback Group. After completing the BIS, participants underwent a similar Go/NoGo task to the one described in Experiment 1. Accordingly, each phase comprised 100 Go and 20 NoGo trials (5:1 Go–NoGo ratio). As reported in Experiment 1, all stimuli remained on the screen for 400 ms, and responses produced brief feedback slides consisting of "correct" or "incorrect" that offset after 400 ms. Go responses had to be performed before stimulus offset to be registered as correct. The inter-trial intervals varied randomly between 1200 and 2400 ms to ensure engagement with the task. All subjects completed a brief practice session (six correct Go or NoGo responses) using the same stimuli that comprised the task. This practice session was conducted with the experimenter present to ensure the comprehension of instructions.

In the Familiar condition, participants were instructed to "Go" on green traffic light stimuli as quickly and accurately as possible, and withhold responses to the red traffic light. Next, a cumulative performance feedback manipulation followed, in which we displayed subjects' percent NoGo accuracy scores on the screen. Participants were informed that the percentage score reflected their performance thus far (they were not informed that the score only reflected NoGo accuracy), and in the next phase of the task,

the Go and NoGo signals would be reversed, such that they would need to make a response as quickly and accurately as possible to the red traffic light, and refrain from responding to the green traffic light. Identical feedback and task instructions were provided to the participants in the Novel condition regarding the change in contingencies of the purple–Go and blue–NoGo associations. It should be reiterated that Experiment 1's results suggest differential impairments across Familiar and Novel conditions regardless of the order in which phases were completed. Therefore, unlike Experiment 1, the phase orders in Experiment 2 were not counterbalanced, in that all participants in the Familiar condition underwent the congruent (Green–Go, Red–NoGo) mappings first, followed by the incongruent mappings; all participants in the Novel condition underwent the Purple–Go, Blue–NoGo mapping first, and these mappings were reversed in the second phase. This change in experimental protocol enabled rendering the congruent contingency as baseline for participants in the Familiar group, and testing whether the presence of a mid-experiment performance manipulation affected subsequent incongruent task performance. An exit survey consisting of demographic questions concluded the experiment (see Figure 11 for a schematic of the experimental design).

No Feedback Group. Participants in the No Feedback group underwent the same procedures as the Feedback group, except that no cumulative performance feedback was provided at any point. This No Feedback group served as a control condition for the Feedback group, as well as an internal replication of Experiment 1.

Data analysis

To examine the role of Feedback, mixed-design ANOVAs with NoGo accuracy as DV, Feedback as a between-subjects and Mapping as a within-subjects factor were

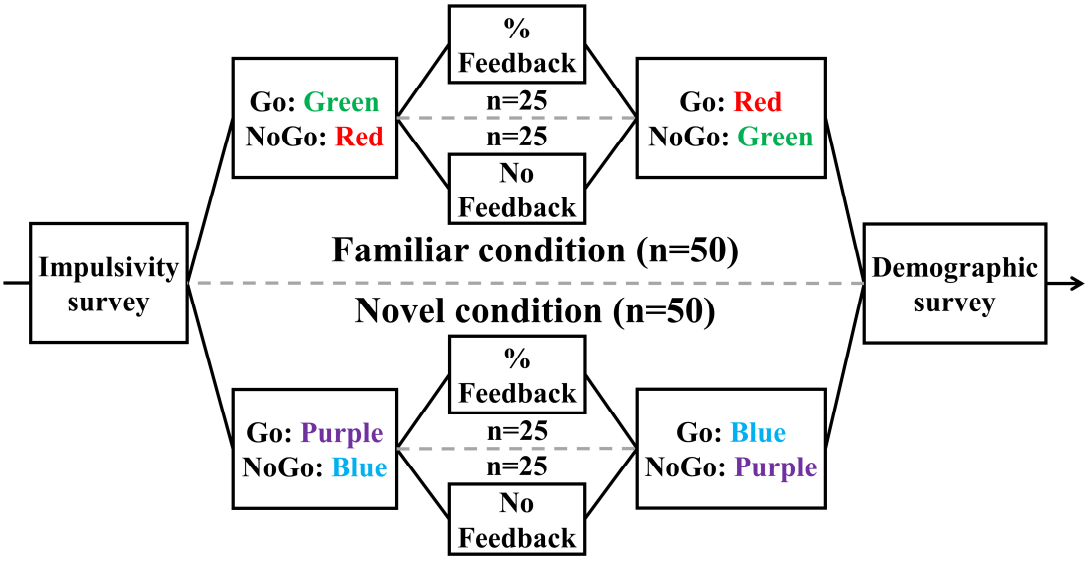


Figure 11. Experiment 2 design. Participants complete an impulsivity survey, then undergo either the Familiar or the Novel conditions of the Go/NoGo task. Similar to Experiment 1, participants perform both color-response mapping phases; however, half of the participants in each condition receive cumulative performance feedback following the first color-response mapping phase. A demographic survey concludes the experiment. %: performance feedback.

performed for each Condition, using the controlled variables Age, Gender, and Impulsivity as covariates. Post-hoc t-tests were carried out to examine mapping-related accuracy differences in both Feedback groups. As a secondary measure of outcome-sensitivity, identical analyses were performed using Go accuracy as a dependent variable.

Building from Experiment 1, we performed a confirmatory omnibus hierarchical multiple regression to test the predictive strength of the Condition and Feedback variables on mapping-related impairment. The summary of the omnibus regression test is reported below, and its details can be found in the Appendix (Supplemental Tables 4 and 5).

We performed a power analysis using the effect size of the Condition x Mapping interaction in Experiment 1 ($\eta_p^2 = .16$) and determined that a sample of 12 participants per group would be sufficient to reach 80% statistical power to detect the effect of differential accuracy rates due to Condition. We opted for this interaction value for our

investigation of the role of feedback, because we wanted our feedback-related assertions to be grounded in predictions of a replicated effect of habitual performance to familiar, and goal-directed performance to novel stimuli. To further increase statistical power due to the addition of a Feedback group per condition, we increased our sample size to 25 per group—a total of 100 undergraduate students.

Results

Primary index of outcome-sensitivity: NoGo accuracy

We hypothesized that performance feedback may be a salient factor that can potentially restore goal-directed control when managing well-established associations. However, cumulative performance feedback did not break the habits elicited by these familiar stimuli. We performed a mixed-design ANOVA using NoGo accuracy as DV, and Age, Gender, and Impulsivity as covariates. We found no main effect of Feedback, $F(1,45) = 0.08$, $p = .778$, $\eta_p^2 < .01$, or Mapping, $F(1,45) = 1.96$, $p = .169$, $\eta_p^2 = .04$, and we also found that no significant Feedback x Mapping interaction exists : $F(1,45) = 0.08$, $p = .776$, $\eta_p^2 < .01$ (see Figure 12; for corroborating regressions, see Appendix,

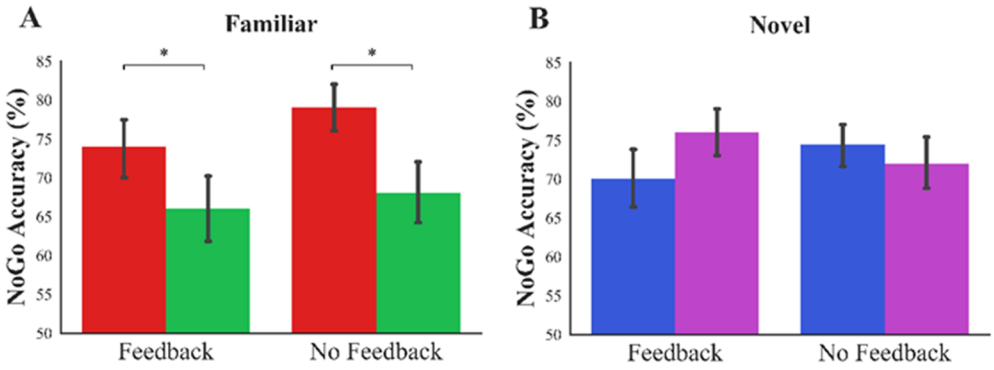


Figure 12. Performance feedback does not significantly disrupt well-established habits. (A) In the Familiar condition, both Feedback and No Feedback groups suffer an incongruency-related impairment ($p = .776$) in NoGo accuracy. (B) NoGo accuracy in the Novel condition is not significantly improved by performance feedback (sig. interaction of $p = .033$, non-sig. post-hoc t-tests: $p > .05$). Error bars denote SEM. Color of bars reflects NoGo stimulus colors.

Supplemental Table 4). Post-hoc t-tests revealed significant incongruency-related impairments in both Feedback, $t(24) = 2.72$, $p = .012$, and No Feedback, $t(24) = 3.16$, $p = .004$, groups, indicating that cumulative performance feedback did not prevent habitual control from dominating in the Familiar condition. Although we were unable to break habits as hypothesized here, our findings lend support to the rigidity of these well-learned associations that persevere in the face of an otherwise salient motivational manipulation, performance feedback (Deci, 1971; Harackiewlcz, 1979).

We performed a similar ANOVA to determine whether cumulative performance tracking improved goal-directed control of novel associations. As seen in Figure 12, we did not find a main effect of Feedback, $F(1,45) = 0.40$, $p = .528$, $\eta_p^2 < .01$, or Mapping, $F(1,45) = 0.60$, $p = .442$, $\eta_p^2 = .01$, yet found a Feedback x Mapping interaction on NoGo accuracy in the Novel Condition when controlling for Age, Gender, and Impulsivity as covariates: $F(1,45) = 4.84$, $p = .033$, $\eta_p^2 = .10$. In sum, these results suggest that performance feedback alone may not be a salient enough manipulation to restore goal-directed control.

Secondary index of outcome-sensitivity: Go accuracy

We performed a mixed-design ANOVA of the Familiar condition data using Go accuracy as DV, Feedback as a between-, and Mapping as a within-subjects factor, with Age, Gender, and Impulsivity as covariates. We found no significant main effect of Feedback $F(1,45) = 0.10$, $p = .751$, $\eta_p^2 < .01$, or Mapping, $F(1,45) = 0.14$, $p = .705$, $\eta_p^2 < .01$, but found a significant Feedback x Mapping interaction: $F(1,45) = 4.73$, $p = .035$, $\eta_p^2 = .09$ (Figure 13), suggesting that Go accuracy was affected differentially by performance feedback. Post-hoc paired-samples t-tests of Go accuracy across phases yielded evidence

for an incongruency-related impairment in the No-Feedback group, $t(24) = 3.22$, $p = .004$), but not in the Feedback group, $t(24) = 1.14$, $p = .265$. Indeed, the omnibus hierarchical regression model attributes Condition and Feedback regressors a significant role in predicting Go accuracy change ($\beta_{Condition} = -0.32$, $p = .001$, $\beta_{Feedback} = 0.28$, $p = .003$; $\Delta R^2 = .18$; see Appendix, Supplemental Table 5).

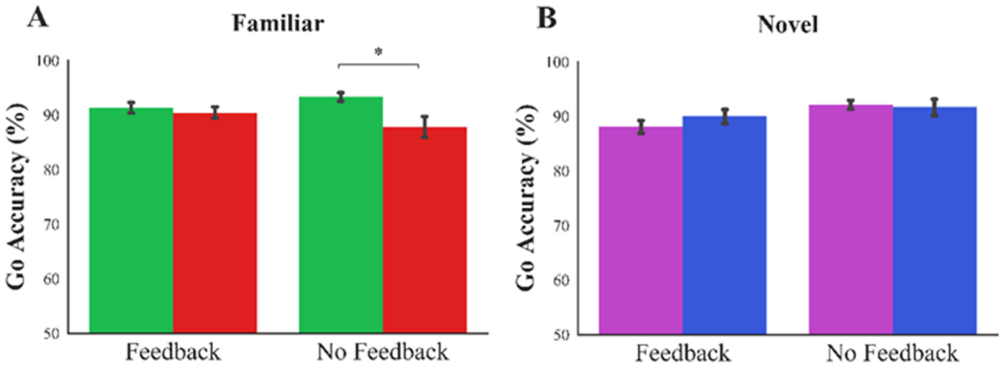


Figure 13. Performance feedback protects against habitual Go actions. (A) When participants received cumulative feedback on their performance, the Go accuracy impairment otherwise observed without feedback was prevented when managing Familiar stimuli (Feedback x Mapping interaction $p = .035$). (B) Performance feedback did not significantly improve Go accuracy in the Novel condition (Feedback x Mapping interaction $p = .117$). Error bars denote SEM. Color of bars reflects Go stimulus colors.

Despite the significant Feedback regressor in the omnibus test, we did not observe a significant improvement effect due to cumulative performance feedback in the Novel condition Go accuracy results. A mixed-design ANOVA using Go accuracy as DV, Feedback as the between-, and Mapping as the within-subjects factor, with Age, Gender, and Impulsivity as covariates revealed no significant main effect of Feedback, $F(1,45) = 3.53$, $p = .067$, $\eta_p^2 = .07$, or Mapping, $F(1,45) = 3.14$, $p = .083$, $\eta_p^2 = .06$, and no significant Feedback x Mapping interaction: $F(1,45) = 2.56$, $p = .117$, $\eta_p^2 = .05$ (Figure 13). Post-hoc paired-samples t-tests suggest an improvement effect only in the Feedback group: $t(24) = -2.39$, $p = .025$ with feedback, $t(24) = 0.32$, $p = .749$ without feedback. Given the lack of significant Feedback x Mapping interaction in the Novel condition, we

refrain from speculating further about the effect of cumulative performance feedback on goal-directed Go responses.

Discussion

In sum, we report that cumulative performance feedback is not sufficient to disrupt the well-learned habits elicited by the familiar stimuli used in our task. However, supplementary analyses using accessory measures of behavioral control (i.e., familiar Go accuracy), suggest that feedback may be a useful tool in enhancing behavioral flexibility. Therefore, these patterns warrant further examination of feedback to disrupt habitual control.

We conclude that cumulative performance feedback was not salient enough to break habits according to our primary analyses, yet our findings were valuable in two ways. First, the validity of our Go/NoGo task using well-learned associations to study habits relies on the rigidity of these green–go and red–stop associations. The persistent habitual control exhibited here despite the delivery of performance feedback lends credence to the associative strength of our familiar stimuli. Next, given the modest signs of performance improvement due to the presentation of performance information, early reports of combined (i.e., performance tracking and monetary incentives) feedback's positive effects on performance, and the beneficial effects of performance-contingent feedback on behavioral flexibility (Kleinsorge and Rinkenauer, 2012; Montague and Webber, 1965; Shen and Chun, 2011; Yee et al., 2016; Zedelius et al., 2012), we were motivated to enhance the salience of the provided feedback to break well-learned habits. In Experiment 3, we further amplified the salience of the outcome by pairing performance-contingent cumulative feedback with a bonus monetary reward prior to

changing Go and NoGo contingencies. We studied the effects of monetary and cumulative performance feedback on Go/NoGo task performance, and whether this amplification of outcome salience resulted in the breaking of a well-learned habit, and improvement of novel, goal-directed performance.

Experiment 3

Methods

Participants

To test the effects of dual feedback, we recruited the same number of participants for Experiment 3 as in Experiment 2. One-hundred participants (76 female, 24 male participants; M_{age}=19.74, SD_{age}=2.79) from the Rutgers University-Newark undergraduate research subject pool were recruited for course credit. All participants provided informed consent. Study protocols were approved by the Rutgers University Institutional Review Board. Participants were excluded if they reported having color-blindness.

Procedures

The promising but insufficient effect of cumulative performance feedback on the motivational control of action motivated us to examine the combined effect of performance and monetary input. Thus, we implemented in our mid-experiment performance feedback manipulation a cash bonus. Experimental procedures were identical to those described in Experiment 2, with the addition of awarding participants in the Feedback group a surprise $5 cash bonus before the change in Go/NoGo contingencies.

After completing BIS, participants underwent a similar Go/NoGo task to the one described in Experiment 2, where they were randomly assigned to Feedback and No Feedback groups, and Familiar and Novel conditions. As in Experiment 2, each phase comprised 100 Go and 20 NoGo trials (5:1 Go–NoGo ratio), and the stimuli remained on the screen for 400 ms. Go and NoGo responses (or lack thereof) produced brief feedback slides consisting of "correct" or "incorrect" that offset after 400 ms. Go responses had to be performed before stimulus offset to be registered as correct. The inter-trial intervals varied randomly between 1200 and 2400 ms to ensure engagement with the task. All participants completed a brief practice session prior to the task, similar to the previous two experiments.

Identical to Experiment 2, in the Familiar condition's first phase, participants were instructed to "Go" on green traffic light stimuli as quickly and accurately as possible, and "NoGo" on red traffic light stimuli. Next, a monetary and cumulative performance feedback manipulation followed, in which we displayed participants' cumulative NoGo accuracy as a percentage score on the screen. Participants were informed that the percentage score reflected their performance thus far. Additionally, unique to Experiment 3, the experimenter left the room, and returned briefly after with a $5 bill, and informed the participant that this money was earned because of performance thus far in the task. Unbeknownst to the participants, the cash bonus was not actually contingent on performance. The participant was then informed that the Go and NoGo signals would be reversed, such that they would need to make a response as quickly and accurately as possible to the red traffic light, and refrain from responding to the green traffic light. Identical performance and monetary feedback information and reversal

instructions were provided to the participants in the Novel condition regarding the reversal of purple–Go and blue–NoGo responses. An exit survey containing demographic questions concluded the experiment (see Figure 14 for a schematic of the experimental design).

Participants in the No Feedback group underwent the same procedures as the Feedback group, except for the feedback manipulation, in that participants received no cumulative performance or monetary feedback.

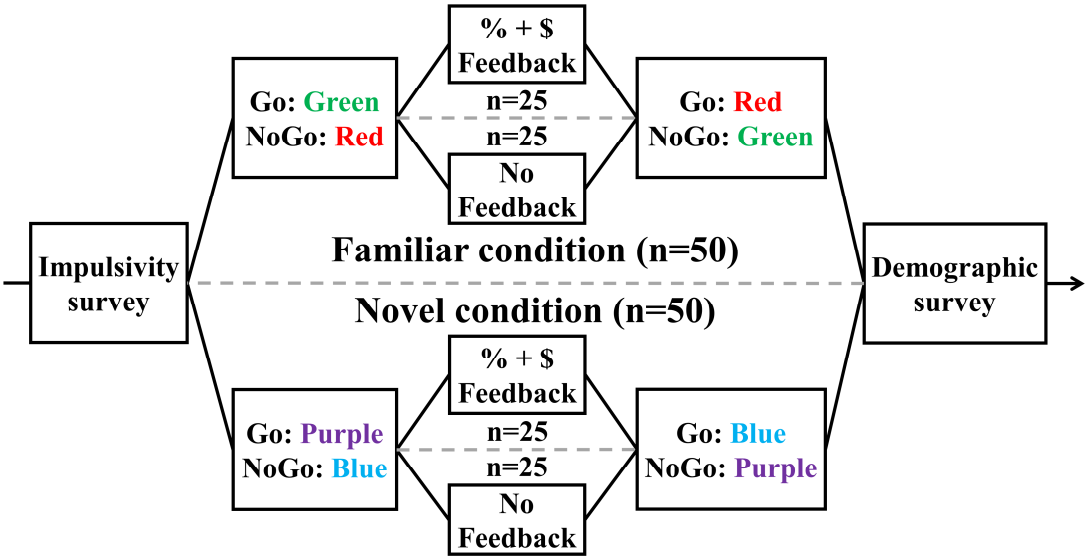


Figure 14. Experiment 3 design. Participants complete an impulsivity survey, then undergo either the Familiar or the Novel conditions of the Go/NoGo task. Similar to Experiment 1, participants perform both color-response mapping phases; however, half of the participants in each condition receive cumulative performance feedback paired with a $5 bonus following the first color-response mapping phase. A demographic survey concludes the experiment. % + $: performance and monetary feedback).

Data analysis

To reveal the potential effect of dual feedback on motivational control, we performed mixed-design ANOVAs with NoGo accuracy as DV, Feedback as a between- and Mapping as a within-subjects factor for each Condition, using the Age, Gender, and Impulsivity variables as covariates. Post-hoc paired-samples t-tests were carried out when necessary to examine mapping-related accuracy differences in both Feedback groups. As

a supplemental measure of outcome-sensitivity, identical tests were performed using Go accuracy as DV. Identical to Experiment 2, we performed a confirmatory omnibus hierarchical multiple regression to test the predictive strength of the Condition and Feedback variables on outcome-sensitivity. The summary of the omnibus regression test are reported below, and the details can be found in the Appendix (Supplemental Tables 6 and 7).

Results

Primary index of outcome-sensitivity: NoGo accuracy

We tested the role of dual feedback in disrupting habitual control to familiar stimuli by performing a mixed-design repeated measures ANOVA on data from the Familiar condition, using NoGo accuracy as the dependent variable. We found no main effect of Feedback, $F(1,45) = 0.75$, $p = .390$, $\eta_p^2 = .10$, or Mapping, $F(1,45) = 1.51$, $p = .225$, $\eta_p^2 = .03$, but found a significant Feedback x Mapping interaction when controlling for Age, Gender, and Impulsivity: $F(1,45) = 5.24$, $p = .027$, $\eta_p^2 = .10$ (see Figure 15). This interaction suggests differential impairment based on the availability of cumulative

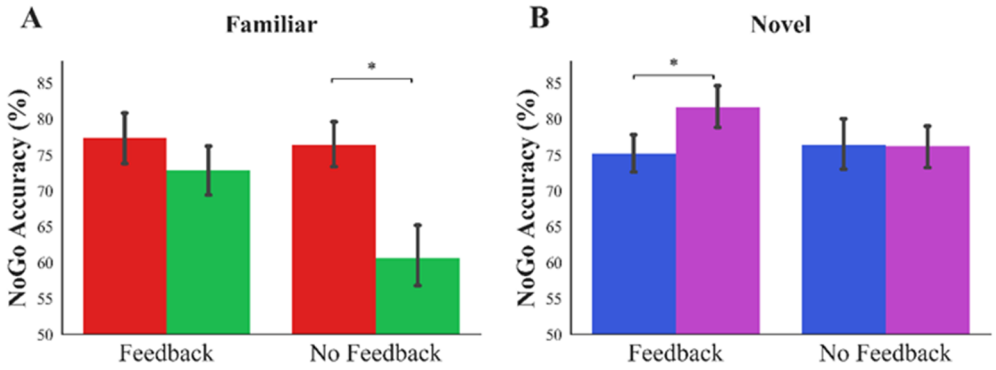


Figure 15. Monetary and performance feedback disrupt habits while improving goal-directed performance to newly-learned stimuli. (A) Providing performance and monetary feedback prevents the incongruency-related impairment normally indicative of habitual control (Feedback x Mapping interaction: $p = .027$). (B) Dual feedback also improves goal-directed control of novel associations significantly (Feedback x Mapping interaction: $p = .038$). Error bars denote SEM. Color of bars reflects NoGo stimulus colors.

performance and monetary feedback, such that the lack of feedback when managing familiar stimuli resulted in a significantly larger incongruency-related decrement in NoGo accuracy. Post-hoc t-tests confirmed a significant impairment in the No-Feedback group, $t(24) = 5.25$, $p < .001$, replicating our findings from Experiments 1 and 2, but no significant effect in the Feedback group $t(24) = 1.92$, $p = .067$.

To understand whether dual feedback enhanced goal-directed performance to newly-learned associations, we performed similar analyses on the Novel condition data. The mixed-design ANOVA, when controlling for Age, Gender, and Impulsivity as covariates, yielded no main effect of Feedback, $F(1,45) = 0.10$, $p = .756$, $\eta_p^2 < .01$, or Mapping, $F(1,45) = 0.42$, $p = .522$, $\eta_p^2 = .01$; however, we found a significant Feedback x Mapping interaction on NoGo accuracy in the Novel condition: $F(1,45) = 4.55$, $p = .038$, $\eta_p^2 = .09$ (Figure 15). Post-hoc t-tests revealed significant improvement of NoGo accuracy in the Feedback group, $t(24) = -2.32$, $p = .029$, which was not observed in the No-Feedback group, $t(24) = 0.08$, $p = .938$.

Consistent with these significant Feedback x Mapping interactions in both Familiar and Novel conditions, our omnibus hierarchical regression model revealed Condition and Feedback regressors to be significant predictors of outcome-sensitivity. Combined, Condition and Feedback explained 26.6% of the variance in mapping-related NoGo accuracy change ($\beta_{Condition} = -0.43$, $p < .001$, $\beta_{Feedback} = 0.28$, $p = .003$; $\Delta R^2 = .27$). These data suggest that the differential mapping-related NoGo impairment observed in Experiment 2 was replicated in Experiment 3, and importantly, that dual feedback is able to significantly predict improvements in performance. The entirety of the omnibus test can be found in the Appendix (Supplemental Table 6).

Secondary index of outcome-sensitivity: Go accuracy

As a supplementary assay of behavioral control, we analyzed Go accuracy using similar statistical procedures. We input Go accuracy as a dependent variable, Feedback as a between-, and Mapping as a within-subjects factor, with Age, Gender, and Impulsivity as covariates into a mixed-design ANOVA. For the Familiar condition, we found no significant main effect of Feedback $F(1,45) = 2.36$, $p = .131$, $\eta_p^2 = .05$, a significant main effect of Mapping, $F(1,45) = 4.15$, $p = .048$, $\eta_p^2 = .08$, but no significant Feedback x Mapping interaction: $F(1,45) = 2.52$, $p = .119$, $\eta_p^2 = .05$ (Figure 16), suggesting that Go accuracy was not significantly affected by dual feedback in the Familiar condition. However, post-hoc paired-samples t-tests revealed incongruency-related impairments in Go actions specific to the No Feedback group: $t(24) = 2.58$, $p = .017$ without feedback vs. $t(24) = 0.10$, $p = .925$ with dual feedback. Given the lack of interaction, we refrain from asserting that dual feedback disrupts habitual Go actions—our secondary assay of outcome-sensitivity.

We then tested the effect of dual feedback on Go accuracy in the Novel condition to determine whether our enhanced feedback manipulation improved goal-directed

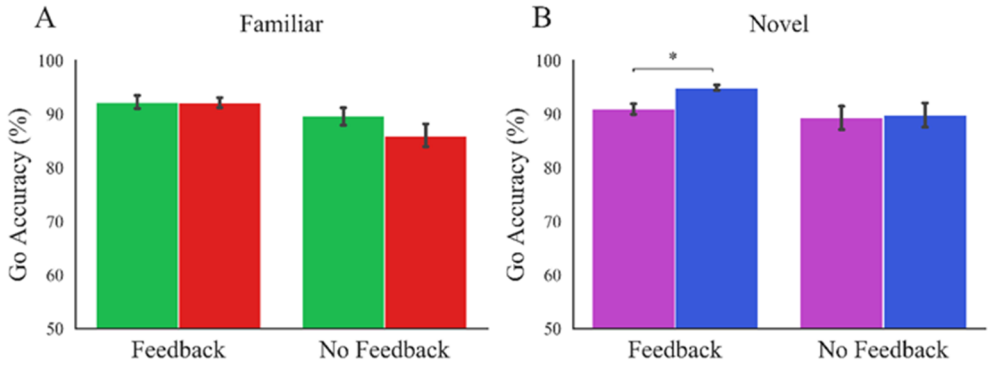


Figure 16. Dual feedback improves goal-directed Go accuracy. (A) Dual feedback did not have a significant effect on the incongruency-related Go accuracy impairment when managing well-learned cues ($p = .119$). (B) Dual feedback improved goal-directed Go responses to novel associations ($p = .012$). Error bars denote SEM. Color of bars reflects Go stimulus colors.

control when managing the contingency changes in newly-learned associations. We performed a mixed-design repeated measures ANOVA using Go accuracy as DV, Feedback as the between-, and Mapping as the within-subjects factor, with Age, Gender, and Impulsivity as covariates. This ANOVA yielded a significant main effect of Feedback, $F(1,45) = 5.49$, $p = .024$, $\eta_p^2 = .11$, and no significant effect of Mapping, $F(1,45) = 0.49$, $p = .488$, $\eta_p^2 = .01$; however, it revealed a significant Feedback x Mapping interaction: $F(1,45) = 6.93$, $p = .012$, $\eta_p^2 = .13$ (see Figure 16). Post-hoc t-tests of each Feedback group confirms that monetary incentives paired with cumulative performance feedback significantly improved newly-learned Go associations that are executed by the goal-directed system: $t(24) = -4.86$, $p < .001$ with dual feedback, $t(24) = -0.51$, $p = .616$ with no feedback.

Our omnibus hierarchical regression model reveals that Condition and Feedback regressors significantly predict mapping-related Go accuracy changes. These regressors in sum account for 21% of the variance in the DV ($\beta_{Condition} = -.36$, $p < .001$, $\beta_{Feedback} = .28$, $p = .004$; $\Delta R^2 = .21$). These values suggest that Go accuracy is selectively impaired in the Familiar condition, and Feedback is able to promote goal-directed Go actions. Due to the non-significant Condition x Mapping interaction in the Familiar condition data, we restrict the scope of our dual feedback assertions on Go accuracy to the Novel condition. Details of the omnibus regression can be found in the Appendix (Supplemental Table 7).

Discussion

Collectively, our Experiment 3 findings suggest that a global motivational boost involving amplified performance and monetary feedback produces a habit-breaking effect that restores goal-directed control. Without feedback, we observe a significant

impairment in NoGo and Go accuracy when familiar green and red light stimuli demand responses incongruent with daily experiences. We find that this outcome-insensitive habit (i.e., inflexible, cue-driven behavior that persists despite the outcome) of the green–go and red–stop actions is disrupted when participants are provided dual feedback, such that the significant incongruency-related NoGo impairment otherwise seen without feedback is prevented. Moreover, our dual feedback manipulation also improves goal-directed control when managing newly-learned associations, as evidenced by significant enhancements to NoGo and Go performance in the Novel group. Possibly, cumulative performance feedback may be enhancing intrinsic motivation. The percentage score may provide individuals the opportunity to track task performance improvements, potentially boosting motivation to improve task-competence (Ryan and Deci, 2000). Paired with the extrinsic reward of a monetary bonus, the dual feedback provided in our experiment may be producing a global increase in motivation, resulting in more deliberate control of otherwise inflexible behaviors.

Importantly, the beneficial effect of such feedback generalizes to more flexible goal-directed performance, as we observe a significant improvement in NoGo and Go accuracy scores to novel blue–go and purple–stop contingencies when participants are provided dual feedback. Without feedback, we find no mapping-related difference in accuracy to novel stimuli, serving as support for the flexible nature of these newly-learned associations that can readily be reassigned per changes in one's environment. These findings identify dual feedback as a powerful predictor of motivational control enhancement.

General Discussion

In a three-experiment study, we introduce a novel Go/NoGo task that capitalizes on familiar, well stamped-in associations of red–stop and green–go to elicit habitual control, and establish dual feedback (i.e., monetary reward paired with cumulative performance tracking) as an effective basis for intervention to break these well-learned habits to restore goal-directed control. We also report enhanced goal-directed, novel learning due to dual feedback, lending support to the effectiveness and scope of our performance enhancing feedback manipulation.

Accordingly, an important goal of our study was to establish our paradigm as a tool that captures real-world habits. In Experiment 1, we demonstrated the rigidity of the familiar green–go and red–stop contingencies compared to the newly-learned, flexible associations. The outcome-insensitive responses elicited by the familiar stimuli were reflected by a significant mapping-related impairment not observed when participants managed novel stimuli. We then tested the strength of the habits evoked in our paradigm by introducing a motivation-based intervention: cumulative performance feedback. This type of feedback was not successful in preventing habitual control, supporting the notion that these existing habits are rigid enough to prevail even in the face of a motivational intervention. Nonetheless, performance feedback was able to produce promising results via secondary assays of behavioral flexibility. Namely, the prevention of habitual "Go" actions motivated the augmentation of our feedback manipulation to amplify its effect on motivational control. In Experiment 3, our combined delivery of performance and monetary feedback prevented the mapping-related impairment that is the result of a habit-dominated action control system, while significantly improving goal-directed control. In

sum, we demonstrated well-existing habits, tested the limits of their associative strength, and provided the foundation for better understanding the restoration of goal-directed control.

Many habit paradigms that emulate the outcome-insensitive nature of habits have in common a shortcoming that limits generalizability to the typical habit experience: difficulty capturing well-learned habits in the lab that can provide a platform for studying habit disruption. Habit strength is limited by the participants' brief exposure to experimental paradigms, and targeting these behaviors that are rendered inflexible in the lab may not be representative of habits encountered in the real world (Ceceli and Tricomi, 2018). Perhaps due to these difficulties, well-learned habits and habit disruption research have been relatively better-represented in field experiments compared to the laboratory setting. For example, several field studies have examined the efficacy of interventions to change various presentations of daily habits , such as recycling and snacking habits (Adriaanse et al., 2009, 2011; Holland et al., 2006). However, recent efforts to bridge lab and field experiments have shown promising results. Although not an experiment of habit disruption, in a recent report, the slips-of-action task in the lab was examined alongside a more ecologically-relevant representation of habits—namely the habit of using one's house keys. In this study, participants demonstrated an outcome-insensitive habit by making key choice errors, such that they persisted in choosing the incorrect key following a change in key covers. The attentional underpinnings of this behavior significantly correlated with slips of action performance, underlining the importance of focusing on well-established behaviors for an improved empirical approach to habit research (Linnebank et al., 2018).

One strategy that has proven beneficial in tackling habit change is implementation intentions, which provides individuals with an if-then plan (i.e., "if X happens, I will do Y"; or in a lab task, "if I see stimulus X, I will press Y")—an aide to override unwanted or inflexible behaviors (Gollwitzer, 1999). In the lab, implementation intentions have produced promising results, albeit with limited efficacy in disrupting strong habits. For instance, Webb and colleagues trained participants for five days on a target detection task, and successfully disrupted this lab-automated association using implementation intentions. However, this planning strategy did not break unwanted smoking habits, lending credence to the idea that the experimental resources at our disposal may not be sufficient in effectively stopping well-established habits (Webb et al., 2009). Although this study approached habitual control from an attentional rather than a value-driven perspective, parallel evidence from the motivational control literature has recently been reported. In another lab study, Verhoeven et al. employed planning strategies within a single experimental session to reduce action slips in an outcome-devaluation task (Verhoeven et al., 2017). Implementation intentions were more effective than goal-intentions (an outcome-based planning strategy, such as "I will not press for outcome X") in reducing action slips when managing abstract images as outcomes, suggesting that implementation intentions may serve as a promising strategy in studying habit disruption—however, effective paradigms to demonstrate well-learned, outcome-insensitive habits, and an intervention to disrupt them are needed. In our study, we developed a task that allowed us to directly capture ecologically significant, well-established habits via the familiar green–go and red–stop associations. We present our Go/NoGo task with familiar and novel stimuli as a strong candidate for demonstrating

habitual behaviors—bridging the success of field studies with the rigor and controllability of lab experimentation. We also illustrate that a salient feedback-based intervention may be utilized to shift cue-driven performance to become value-driven, laying the foundation to translational applications.

Our work also asserts that the use of familiar stimuli may circumvent the obstacles of training length and stimulus–response strength in habit research—an important step in improving paradigms to foster effective habit disruption strategies. A few prior studies have considered a similar approach. In a study investigating habits in substance use disorder, McKim and colleagues induced stimulus familiarity by pre-training a set of stimuli, and tested the strength of the familiar versus novel stimulus sets on a subsequent day via the reversal of a sub-set of these contingencies (McKim et al., 2016). They found that compared to healthy controls, individuals with substance use disorder performed better in well-learned stimulus–response execution, yet exhibited impairments in managing contingency reversal. In cadence with these findings, our study reveals that when managing contingencies that have been well-established throughout development—beyond an experimental pre-training stage—the recruitment of the habit system may also be evident in healthy individuals. Similarly, developmental and clinical researchers have used familiar green and red stimuli in Go/NoGo tasks with children suffering from attention-deficit/hyperactivity disorder, as well as healthy adults to reduce task demands, and justified their decision by identifying these colors as having developmental relevance (Mostofsky et al., 2003; Suskauer et al., 2008). These prior reports highlight the utility of capitalizing on existing associations when examining habits, especially for clinical examinations of behavioral rigidity. The current report

contributes further by introducing a task that requires minimal familiarity training, and by the inclusion of a motivational strategy to disrupt the familiarity-driven outcome-insensitivity. These contributions may be especially useful for optimizing costly fMRI designs, and benefit future translational neuroscience work that aims to reveal the neural bases of habit disruption.

The science of habits is a domain with direct clinical applications. The treatment of habit-based pathologies (e.g., obsessive-compulsive disorder) are within the scope of the habit literature, yet our field's disproportionate focus on the formation of rigid behaviors, rather than overcoming well-formed habits, limits the translational impact of our research (Griffiths et al., 2014). Indeed, several studies have highlighted the habitual aspects of various clinical disorders, as well as their underlying neural mechanisms (e.g., Alvares et al., 2014, 2016; Banca et al., 2015; Delorme et al., 2016; Gillan et al., 2015; McKim et al., 2016; Morris et al., 2015; Reiter et al., 2016; Sjoerds et al., 2013). Researchers have further employed neurotransmitter depletion to emulate the biochemical profiles of psychopathologies to detect action control deficits (de Wit et al., 2012; Worbe et al., 2015, 2016). Sub-clinical symptom presentation has also been investigated from the perspective of action control (Dietrich et al., 2016; Hogarth et al., 2012; Morris et al., 2017; Snorrason et al., 2016). Furthermore, the multi-faceted role of stress in dictating motivated behaviors has been extensively demonstrated under acute, chronic, interaction of acute and chronic, and pharmacologically induced stress hormone reactivity (Radenbach et al., 2015; Schwabe et al., 2008, 2012; Schwabe and Wolf, 2009, 2010, 2011; Soares et al., 2012; Taylor et al., 2014). Therefore, although researchers have characterized numerous contexts in which habits are prevalent, interventions that restore

goal-directed motivational control have not been examined with similar vigor. As we demonstrate the habit-breaking effects of pairing monetary reward with cumulative performance feedback to amplify the salience of goals, we highlight the need for research avenues that not only identify goal-directed control deficits in clinical disorders, but work toward restoring these deficits to improve treatment strategies and quality of life.

Conclusions

The disproportionate focus on habit formation and expression in the literature motivated us to direct our efforts to an area of habit research less-explored: habit disruption. Although much research now confirms the habitual aspects of various pathologies, studies examining the restoration of these behavioral rigidities are relatively scarce. Here, we introduce a task that allows us to examine a more complete signature of motivational control by capturing well-learned habits and newly-learned goal-directed behaviors, as well as the possibility to test manipulations that may restore deliberate control. This method may be especially beneficial for understanding the neural markers of motivational control in healthy and compromised populations, as it capitalizes on existing associations that do not require extended lab-training. We also underline the efficacy of feedback in disrupting well-learned habits and promoting outcome-driven, goal-directed behaviors. This motivation-based manipulation may further inform the mechanisms underlying the habit disruption process—a translationally valuable research domain with direct clinical relevance.

Chapter 4: Investigating well-learned habits and their disruption as a function of ADHD symptom severity (submitted)

Introduction

Individuals with attention deficit-hyperactivity disorder (ADHD) are known to exhibit cognitive impairments that span domains of attention and impulsivity (American Psychiatric Association, 2013). These hallmark symptoms are often accompanied by executive control irregularities, such as diminished inhibitory control and excessive distractibility that interfere with daily functioning (Willcutt et al., 2005). Additionally, behavioral and neurobiological reports have highlighted reward-related abnormalities in ADHD, in that individuals with ADHD display impairments in learning from, interacting with, and processing rewards (Castellanos and Tannock, 2002). Children and adults with ADHD present heightened delay aversion, such that they choose immediate, less valuable rewards over delayed yet larger rewards (Antrop et al., 2006; Kessler et al., 2005a; Marx et al., 2013; Sonuga-Barke et al., 1992). In addition to such examples of suboptimal decision-making, individuals with ADHD also exhibit abnormal reward-related neural processing in the brain's reward circuitry, such as decreased signaling in the ventral striatum during reward anticipation, and atypical orbitofrontal cortex (OFC) activity during reward delivery (Furukawa et al., 2014; Plichta and Scheres, 2014; Ströhle et al., 2008; von Rhein et al., 2015; Wilbertz et al., 2012). The affected regions of the brain that regulate reward anticipation and processing (i.e., the striatum and prefrontal cortex), are also known as integral areas for executing motivated behaviors (Balleine and O'Doherty, 2009; O'Doherty, 2016). These neurobehavioral dysfunctions in ADHD, when taken together with the cardinal presentations of inattention and impulsivity, suggest potential disparities in the control of motivated behaviors that have yet to be elucidated.

The motivational account of behavioral control posits that our actions can be either goal-directed, as in, performed deliberately in pursuit of a desirable outcome, or habitual, as in, triggered in response to a salient cue regardless of outcome value (Dickinson and Balleine, 1994). These components of motivational control have distinct neural signatures, such that the prefrontal cortex and caudate are known to be imperative for the execution of goal-directed behaviors, while cue-based habitual control is largely associated with the putamen and motor cortex (Haber, 2003; O'Doherty et al., 2004; Tricomi et al., 2009). Interestingly, a compelling body of work documents functional and structural abnormalities in ADHD when compared to neurotypicals (NTs) in these brain regions, suggesting a compromised corticostriatal system that could be indicative of motivational control deficits. For example, ADHD is associated with reduced gray matter volume in the caudate, expansion of the posterior putamen, and aberrant connectivity in the ventromedial prefrontal cortex (vmPFC) and anterior cingulate cortex (ACC) (Costa Dias et al., 2013; Frodl and Skokauskas, 2012; Norman et al., 2016; Qiu et al., 2009; Rosch et al., 2018; von Rhein et al., 2017). Studies in rodents have suggested that a rat model of ADHD, the spontaneously hypertensive rat, exhibits a habit-dominated motivational control system, in that these rats that possess ADHD-like symptoms also display outcome-insensitive behavioral patterns (i.e., pressing a lever that predicts a food outcome to which the rat is sated) (Natsheh and Shiflett, 2015). Neural evidence suggests that this behavioral deficit is linked to imbalances in dopamine receptor activation, supporting the idea that abnormalities in the striatal systems may also manifest as an over-reliance on habitual control in ADHD (Natsheh and Shiflett, 2018).

If ADHD is indeed associated with enhanced habitual control that favors outcome-insensitive behaviors, the next logical and translationally valuable step would be to identify strategies that can overcome this behavioral deficit. For instance, performance-contingent feedback is a frequently employed tool that has been shown to improve behavioral output (Kluger and DeNisi, 1996; Montague and Webber, 1965). The positive effects of feedback in the form of performance-tracking information, as well as primary and secondary incentives, have been well-documented in the cognitive flexibility domain—namely using task-switching paradigms. Indeed, even the promise of a future performance-contingent reward has been shown to amplify task-switching performance (Yee et al., 2016). Importantly, performance-contingent monetary feedback is associated with the engagement of top-down control of task-switching processes (Umemoto and Holroyd, 2015). Taken together, we believe that the benefits of feedback on behavioral output and control over actions may carry over to the restoration of goal-directed behaviors in ADHD. In support of this hypothesis, we have previously demonstrated the beneficial effects of feedback on the motivational control of action (Ceceli et al., submitted).

Tackling the expression of habits and the restoration of goal-directed behaviors in potentially compromised populations may involve overcoming the methodological limitations of the traditional habit paradigm. A meaningful assessment of habit expression and disruption may require access to rigid habits with a strong association between the triggering stimulus and the behavioral response. Therefore, instead of relying on labile, newly-learned habits that have been the subject of inquiry in most investigations of motivational control (Ceceli and Tricomi, 2018), it may be more

effective to study habit expression and disruption via well-learned, existing S–R associations that do not require extensive training in the laboratory (Ceceli et al., submitted).

To this end, we developed a Go/NoGo task that capitalizes on familiar green and red traffic light stimuli that activate existing stimulus–response associations (Ceceli et al., submitted). If green-Go and red-NoGo associations are habit-driven, an incongruent Go/NoGo mapping (green-NoGo, red-Go) should produce significant decrements in accuracy. Importantly, Go/NoGo mappings that involve novel stimuli with no significant behavioral representations (i.e., blue and purple light stimuli) should evoke no mapping-related performance impairments. If ADHD is associated with heightened habitual control, symptom severity might track the mapping-related impairments elicited by the familiar Go/NoGo stimuli (e.g., higher symptom severity scores should predict heightened errors of commission—response execution when instructed to withhold). Furthermore, if performance and monetary feedback are effective in restoring goal-directed control, this dual feedback delivery should protect against the mapping-related accuracy impairment, preventing the increase in commission errors when Go and NoGo associations are incongruent with daily experiences. Similarly, such a disruption in habits may also be correlated to ADHD symptom severity, such that a more severe presentation of ADHD symptoms may be less affected by the beneficial effects of feedback. Alternatively, if feedback is a salient enough motivator, highly symptomatic individuals may also benefit from our feedback manipulation, resulting in habit disruption across the board. To reveal whether ADHD is associated with habitual control, and whether a habit-dominated motivational control system may be remediated, we administered our well-

learned habit task over the course of two days on a large sample from the general population, from whom we collected ADHD-related symptomology information. On the first day, we examined the execution of well-learned habits in our sample, and on the second day, we introduced our motivational enhancement manipulation—a combined delivery of performance information and monetary feedback—to restore goal-directed control. Importantly, per our pre-registered analysis plan (document URL: https://osf.io/fjcbw), we used ADHD-related measures to detect whether symptoms of the disorder tracked well-learned habit expression and disruption.

Methods

Participants

To determine the sample size for our study, we performed an *a priori* power analysis on data from an existing study that examined inhibitory control capacity and ADHD-related symptoms (Wodushek and Neumann, 2003). In this study, healthy adults were categorized into high vs. low ADHD symptom groups for inhibitory control comparisons. We extracted effect sizes from the correlations between inhibitory control and non-verbal inattention in both symptom severity groups, and averaged the two resulting projected sample sizes. The averaged sample size needed to reach 80% statistical power was determined to be 105. We recruited 106 participants to make up for one participant's corrupted data. Thus, 106 undergraduate students (79 female, 27 male; M_{age} = 20.23, SD_{age} = 4.07) from the Rutgers University-Newark campus participated for course credit. Informed consent was provided by all subjects per Declaration of Helsinki human subject protection guidelines. The Rutgers University Institutional Review Board approved study protocols. Participants were excluded from participation for self-reported

color-blindness. Two participants' data were excluded from analyses due to attrition (n=1) and data corruption (n=1). Thus, the statistical analyses were performed on the remaining 104 participants (77 female, 27 male participants; M_{age} = 20.20, SD_{age} = 4.10).

Materials and procedures

Participants performed Go/NoGo tasks adapted from Ceceli et al. (submitted) over two days. On day one, all participants underwent Go/NoGo tasks with familiar green and red traffic light stimuli (Familiar condition), and novel blue and purple traffic light stimuli (Novel condition) as Go and NoGo signals. Participants were instructed to respond as quickly and accurately to these stimuli as possible using the keyboard. A second phase followed in each Condition (Familiar/Novel), where the color-response mappings were swapped (see Figure 17). In the Familiar condition, the Green-Go/Red–NoGo color-response mapping was considered "congruent" with daily experiences, while the Red–Go/Green–NoGo mapping was

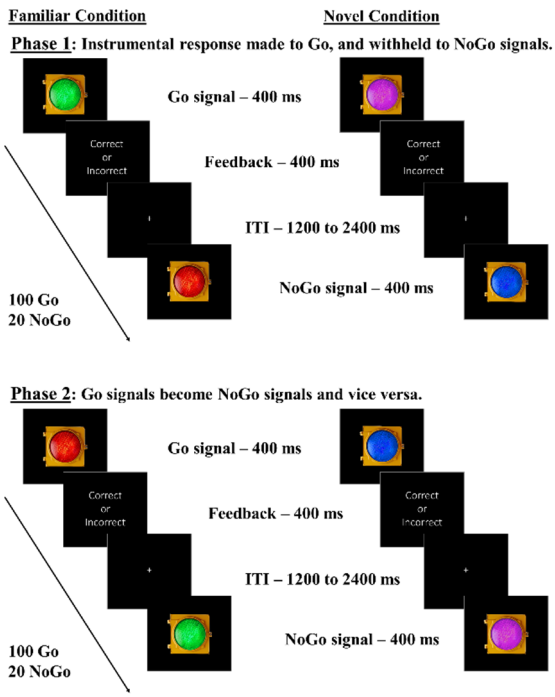


Figure 17. Go/NoGo task with familiar and novel lights. Participants undergo both Familiar and Novel conditions. In the Familiar condition, participants complete two phases: one in which green represents Go and red represents NoGo ("congruent" mapping), and one in which red represents Go and green represents NoGo ("incongruent" mapping). In the Novel condition, participants complete two similar phases, but the colors are blue and purple, for which we assume no strong pre-existing associations with go/stop responses. We predicted more commission errors in the Familiar condition for incongruent than congruent mappings, indicating outcome insensitivity, with no such within-subject differences expected in the Novel condition. Phase and Condition orders were counterbalanced across subjects.

considered “incongruent”, in that it required the participant to override the well-established go and stop meanings of these stimuli. The Novel condition stimuli, however, are assumed to have no well-established Go or NoGo associations in daily life, in that the swapping of the color-response mappings should not require overriding associations that have been well-established. If familiar associations elicit habitual, cue-driven behavioral control, participants should experience a significant impairment in NoGo accuracy when green is mapped with NoGo. In the Novel condition, participants should perform similarly when managing either color-response mapping due to blue and purple not being strongly associated with Go/NoGo signals, reflecting goal-directed performance. We counterbalanced the order in which participants underwent the two phases within each Condition to ensure that our results were not due to a specific order of managing color-response contingencies. We also counterbalanced the order in which participants underwent the Familiar and Novel conditions. Lastly, participants completed the Adult ADHD Self-Report Scale (ASRS), a two-part survey that captures inattentive and hyperactive symptom manifestation associated with ADHD (Kessler et al., 2005b), and a demographic survey, concluding day one’s procedures.

Day two was completed within three days of day one and examined the potential habit-disrupting effect of a motivational enhancement. On day two, all participants underwent the Familiar condition of the Go/NoGo task, completing the “congruent” color-mapping first. Next, we induced motivational enhancement via the delivery of cumulative performance feedback and a monetary incentive. Specifically, participants’ cumulative task performance was displayed as a percentage score on the screen. Additionally, the experimenter briefly left the room, returning shortly after with a $5 cash

bonus. The participants were informed that the $5 bonus was due to their performance on the task. The participants were then instructed to perform the "incongruent" color-mapping of the Familiar condition, and were informed that they may receive another performance-contingent cash bonus afterwards. Unbeknownst to the participants, the mid-session cash bonus was not actually contingent on performance. We did not counterbalance color-mapping of Go/NoGo contingencies on day two to render the congruent color-mapping performance as baseline. Thus, we were able to test whether the presence of a mid-experiment motivational manipulation affected subsequent incongruent color-mapping performance (i.e., overriding the green-Go/red-NoGo habit). Lastly,

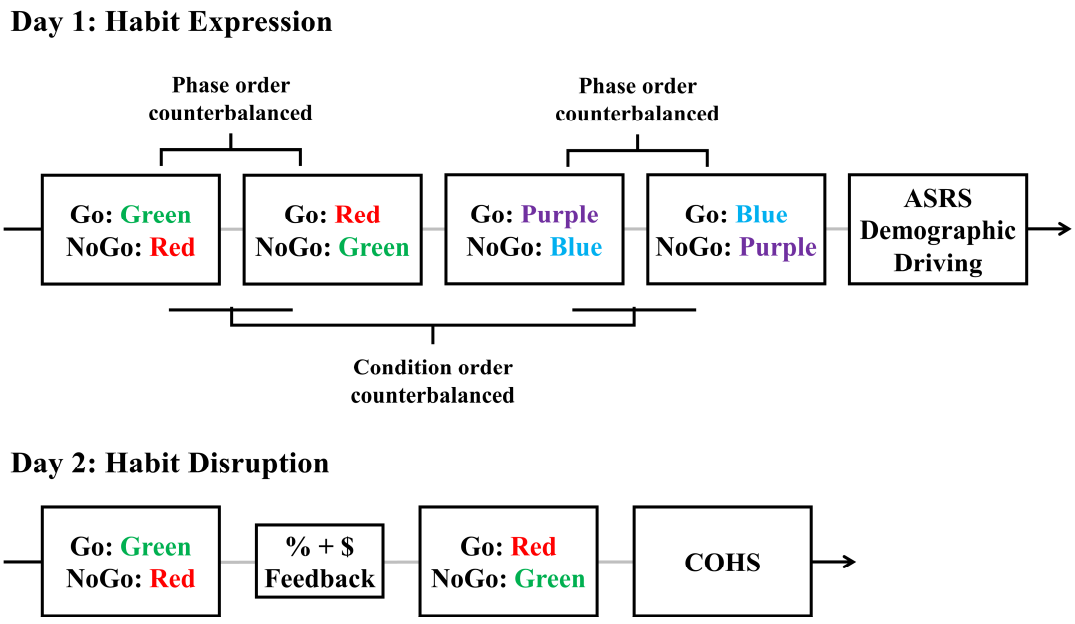


Figure 18. Experimental design. Day 1: all participants undergo the Familiar and the Novel conditions of the Go/NoGo task, completing both color-response mappings within each condition. We counterbalance the order in which the Familiar and Novel conditions, and the color-response mapping phases are managed. Participants complete ADHD, demographic, and driving experience related measures to conclude the session. Day 2: all participants return to complete the Familiar condition of the Go/NoGo task. Participants first manage the color-response contingencies that are congruent with their daily experiences, and are provided cumulative performance feedback paired with a $5 bonus. Participants then complete the color-response contingencies that are incongruent with their daily experiences. Completion of the COHS survey concludes the study.

participants completed the Creature of Habit Survey (COHS) (Ersche, Lim, Ward, Robbins & Stochi, 2017), quantifying the frequency of daily habitual tendencies, and a brief post-experiment questionnaire (see Figure 18 for a schematic of the experimental design).

In each phase, there was a 5:1 Go/NoGo ratio, with 100 Go and 20 NoGo trials. Each Go/NoGo stimulus remained on the screen for 400 ms. Participants were required to respond to Go signals before the offset of the stimulus for a correct response. After offset, each response produced a brief "correct" or "incorrect" text slide. To ensure engagement with the task, inter-trial intervals varied randomly between 1200 and 2400 ms. Participants completed a practice session prior to each Condition, which consisted of six correct Go or NoGo responses using that condition's stimuli. The experimenter remained present to ensure the instructions were understood during the practice sessions.

Data analysis

We pre-registered our task procedures and analyses prior to data collection via the Open Science Framework project registration portal (document URL: https://osf.io/fjcbw, see the Appendix for a copy of the report). Analyses that were not outlined in our pre-registration document are marked as exploratory below. Data analysis was performed using the nlme package in R (version 3.5.1).

We used NoGo accuracy as our primary measure of outcome-sensitivity, as the high NoGo to Go ratio was hypothesized to produce pre-potent Go responses. As a secondary measure of outcome-sensitivity, we also performed all analyses using Go accuracy to supplement our assertions of differential outcome-sensitivity across Familiar and Novel conditions, and reveal the potential role of ADHD symptom severity in

contributing to outcome-sensitivity. Participants with standardized residuals less than -3.3 and greater than 3.3 were identified as outliers (Tabachnick and Fidell, 2007). Analyses excluding outliers are reported if data removal produces substantial changes in results (i.e., changes in statistical significance of any regressor).

ADHD symptom severity and well-learned habits

We performed an omnibus regression test to discern the contributions of symptom severity on outcome-sensitivity within Familiar and Novel condition data collected on day 1. We used ΔNoGo_Accuracy (i.e., change in NoGo accuracy scores across mappings) as our dependent variable (DV) to measure the within-subject mapping-related change in accuracy. A greater mapping-related impairment represents greater outcome-insensitivity (e.g., heightened difficulty overriding a color-response mapping). In a hierarchical structure, we first input the regressors Age, Gender, Condition_Order (order in which participants underwent Familiar and Novel conditions), Phase_Order (order in which participants underwent color-response mappings within each Condition), and Driving (each participant's experience driving, scaled in months), with Subject as a random factor into a linear mixed model. This model extracted the predictive strength of each of these controlled variables on outcome-sensitivity. In the next hierarchical step, we added the regressors ASRS_Inattentive (part A of the ASRS measure capturing symptoms of inattention), ASRS_Hyperactive (part B of the ASRS measure capturing symptoms of hyperactivity), and ASRS_Total (parts A and B aggregated to derive a composite score of ADHD symptom severity). Because our sample included six participants who had received ADHD diagnoses, we also input a Diagnosis regressor to determine whether clinical manifestation of ADHD—albeit in a small proportion of

participants—affects outcome-sensitivity. We used COHS scores as a regressor to find potential correlations with tendency to behave habitually in daily life and outcome-sensitivity in our task. These regressors served to explain the main effects of each individual difference measure on outcome-sensitivity. In the third step of the hierarchical model, we input Condition (Familiar/Novel) as a regressor to specifically detect whether participants exhibited differential outcome-sensitivity across Familiar and Novel conditions. A significant contribution of this variable would confirm that the familiar red and green stimuli indeed elicit outcome-insensitive, habitual control, while the novel stimuli are labile, and thus controlled by goal-directed processes. We performed post-hoc t-tests of NoGo accuracy between phases in each Condition to ascertain differential mapping-related impairment across Familiar and Novel conditions. Lastly, because of our specific focus on the influence of ADHD symptomology on habitual control, we also entered all individual difference measures' interactions with Condition as regressors (e.g., ADHD_Inattentive x Condition) into step four of the model. Thus, we were able to distinguish the effects of each variable on outcome-sensitivity across Familiar and Novel conditions.

In brief, we expected the controlled demographic and counterbalancing variables (Age, Gender, Driving, Condition_Order, and Phase_Order) to be trivial in predicting outcome-sensitivity. We did not expect the Driving regressor to play a significant role in altering outcome-sensitivity, as we expect our well-learned habit task to capture well-established associations that extend beyond experience with these color-response mappings in a traffic context. We input both main effect and interaction regressors related to individual differences in ADHD symptomology and daily habitual tendencies to reveal

potential associations with outcome-sensitivity. This way, we were able to inquire whether these individual difference regressors yielded strong associations with global outcome-sensitivity (i.e., main effects predicting mapping-related impairments independent of stimulus familiarity), and further interrogate whether such an association existed with well-learned habit expression in particular (i.e., ADHD-related measure x Condition interaction predicting an effect on outcome-sensitivity differentially across Familiar/Novel conditions). We also expected Condition to serve as a significant predictor in driving outcome-sensitivity, as the Familiar condition stimuli should selectively elicit outcome-insensitive habits, while the Novel condition stimuli should have no such effect on behavior.

ADHD symptom severity and habit disruption

We have previously shown the habit-disrupting effect of cumulative performance and monetary feedback (Ceceli et al., submitted). Here, we test via another omnibus regression whether ADHD symptom severity predicts habit disruption success. We performed a similar linear mixed model on the aggregate of Familiar data across two days, encompassing performance to the Familiar stimuli with and without feedback. We input our controlled variables of Age, Gender, Driving, Condition_Order, and Phase_Order, with Subject as a random factor into the first step. Our model similarly included ASRS_Inattentive, ASRS_Hyperactive, ASRS_Total, Diagnosis, and COHS in the second step to detect the main effects of individual differences on outcome-sensitivity. In the third step, our regression included a Feedback regressor that coded the availability of the mid-experiment dual-feedback manipulation. Because this analysis was performed only on the Familiar condition data (the Novel condition was not administered

on the second day with feedback), we included no Condition regressor. Lastly, we included in step 4 our individual difference measures' interactions with Feedback as regressors (e.g., ASRS_Inattentive x Feedback) to examine habit disruption per variations in ADHD-related behaviors and daily habitual tendencies.

Similar to our previous omnibus regression, we expected trivial contribution from our controlled variables, but a significant contribution from the Feedback regressor, as the delivery of dual feedback should disrupt the well-learned habit. We expected that symptom severity may affect outcome-sensitivity globally (significant main effects of individual difference measures), but also differentially across Feedback sessions (e.g, significant contribution of ADHD_Inattentive x Feedback). Additionally, we identified an alternative hypothesis—the possibility of habit disruption across the board (pre-registration document, Hypothesis 2b_alt). We expected no directionality in subtypes governing outcome-sensitivity (as in, inattentiveness or hyperactivity specifically driving habits), but we do note that if either subtype plays a major role in driving motivational control in the previous omnibus regression detecting the role of symptom severity on habitual control, that same subtype should predict habit disruption. We expected the frequency of habitual tendencies in daily life, as assayed by COHS, to yield a negative correlation with habit disruption (i.e., a significant COHS x Feedback result).

Supplementary index of outcome-sensitivity: Go accuracy

We used Go accuracy as a supplemental measure of outcome-sensitivity. Thus, we repeated all mixed models that examined ΔNoGo_Accuracy using ΔGo_Accuracy as DV.

Exploratory analyses: Go RT and individual difference measures

We extended our analyses beyond the pre-registered plans and explored the potential correlations between Go reaction time (RT) and our individual difference measures of symptom severity (ASRS_Inattentive and ASRS_Hyperactive) and daily habitual tendencies (COHS). These variables were entered into a correlation matrix, and Pearson's *r* values were corrected for multiple comparisons using the Holm–Bonferroni method. Specifically, we expected a negative correlation between RT and our individual difference measures. Most notably, we expected such an association between RT and ASRS_Hyperactive, which would suggest quicker familiar Go actions to be associated with pronounced hyperactivity.

Results

The summary of our sample's demographic and individual difference measure profile can be found in Table 2. First order correlations involving the primary individual difference measures (i.e., ASRS_Inattentive, ASRS_Hyperactive, COHS) and measures of outcome sensitivity (i.e., ΔNoGo_Accuracy and ΔGo_Accuracy) can be found in the Appendix, Supplemental Table 8.

Table 2. Descriptive statistics of sample profile.

Variable	Mean	SD	Range
Age	20.20	4.10	18–45
Driving experience	31.77	50.05	0–324
ASRS_Inattentive	15.86	5.42	4–28
ASRS_Hyperactive	13.08	5.17	0–25
ASRS_Total	28.77	9.30	5–50
COHS	99.03	13.74	67–131

Note: N = 104 (77 females, 37 males). Age measured in years, driving experience measured in months. SD = standard deviation.

ADHD symptom severity and well-learned habits

We performed a linear mixed model using ΔNoGo_Accuracy as the DV and Subject as a random factor to determine whether ADHD symptom severity significantly predicts outcome-sensitivity in our well-learned habit task (see Table 3). Our proposed model did not meet the assumptions of non-multicollinearity, in that three pairs of fixed factors were highly correlated with each other (for the associated Variance Inflation Factors, see the Appendix, Supplemental Table 9). Thus, we report the analyses as registered in the Supplement, and report below an adjusted model that meets the assumptions of multicollinearity, normality and homoscedasticity. Specifically, we revised our model to remove the regressors Age, Condition_Order, and ASRS_Total to prevent multicollinearity with the regressors Driving, Phase_Order, and ASRS_Inattentive/Hyperactive that are more crucial for our hypotheses.

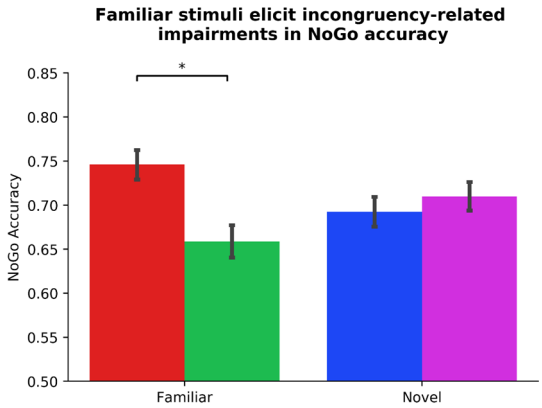


Figure 19. Familiar stimuli elicit incongruency-related impairments in NoGo accuracy. Participants exhibit outcome-insensitivity when managing familiar stimuli with color-response mappings that are incongruent with their daily experiences ($p < .001$). Newly-learned Go/NoGo signals evoke no significant change in NoGo accuracy regardless of color-response mapping, indicating intact goal-directed performance ($p = .279$). The differential habit expression effect across Conditions depicted here is independent from ADHD symptom severity (see Table 3 for individual difference measure contributions to habit expression). Color of bars reflects NoGo stimulus colors.

Standard within-group residuals were within -3.3 and 3.3; thus no participants were identified as outliers (Tabachnick and Fidell, 2007). In the first step of our hierarchical mixed model, contrary to our hypothesis, Gender significantly predicted outcome-sensitivity, β_{Gender} -0.15, $p = .036$, in that female participants displayed

significantly worse mapping-related impairments. Neither Driving experience nor the counterbalancing variable, Phase_Order, predicted outcome-sensitivity (*p*s > .252), model R^2= 0.03. In the second step of the model, we added the individual difference measures of ADHD symptom severity, clinical ADHD diagnosis, and frequency of habitual tendencies in daily life (COHS). We found no main effects of individual difference measures on outcome-sensitivity (all *p*s > .548). The log likelihood estimate derived by comparing first and second steps of our model yielded no significant global (as in, non-Condition specific) contribution attributable to the ASRS_Inattentive, ASRS_Hyperactive, Diagnosis, and COHS regressors, $\chi^2(4) = 0.70$, $p = .952$, R^2= 0.03, $\Delta R^2 < 0.01$. In the third step, we entered the Condition regressor, which significantly improved the predictive strength of the model, $\chi^2(1) = 21.53$, $p < .001$, R^2= 0.13, ΔR^2 = 0.10, $\beta_{Condition} = 0.31$, $t(103) = 4.66$, $p < .001$, meaning outcome-sensitivity was differentially affected by whether participants managed the Familiar or Novel versions of the task. Post-hoc t-tests confirmed that mapping-related NoGo accuracy impairments were evident only when managing Go/NoGo contingencies in the Familiar condition, $t(103) = 5.33$, $p < .001$, while performance in the Novel condition was comparable regardless of color-mapping associations, $t(103) = -1.09$, $p = .279$ (see Figure 19). In the fourth step of the model, we input the interaction of each individual difference regressor with Condition to detect their potentially differential effects on outcome-sensitivity across Familiar and Novel conditions, but found no significant contribution from any ADHD-related or daily habit frequency variable (all *p*s > .085, $\chi^2(4) = 6.19$, $p = .186$, R^2= 0.15, ΔR^2 = 0.03). These results suggest that our sample exhibited outcome-insensitive well-learned habits across the board, but the degree of habitual control as

assessed by change in NoGo accuracy was not significantly related to ADHD symptom severity.

Table 3. Hierarchical Mixed Model of ADHD Symptomology and Habit Expression: ΔNoGo_Accuracy.

Variable	***VIF***	***β***	***t***	***sig.***
Model 1				
Gender	**1.01**	**-.15 (.07)**	**-2.13**	**.036**
Phase_Order	1.01	-.01 (.07)	-0.09	.931
Driving	1.00	.08 (.07)	1.15	.252
Model 2				
Gender	**1.08**	**-.14 (.07)**	**-2.00**	**.049**
Phase_Order	1.04	-.01 (.07)	-0.15	.877
Driving	1.30	.08 (.08)	1.08	.283
ASRS_Inattentive	1.62	-.01 (.09)	-0.08	.939
ASRS_Hyperactive	1.71	.05 (.09)	0.54	.591
Diagnosis	1.30	.01 (.08)	0.14	.891
COHS	1.06	-.04 (.07)	-0.60	.548
Model 3				
Gender	**1.08**	**-.14 (.07)**	**-2.10**	**.039**
Phase_Order	1.04	-.01 (.07)	-0.16	.871
Driving	1.30	.09 (.08)	1.13	.260
ASRS_Inattentive	1.62	-.01 (.08)	-0.08	.936
ASRS_Hyperactive	1.71	.05 (.09)	0.57	.573
Diagnosis	1.30	.01 (.08)	0.14	.885
COHS	1.06	-.04 (.07)	-0.63	.528
Condition	**1**	**.31 (.07)**	**4.66**	**<.001**
Model 4				
Gender	**1.08**	**-.14 (.07)**	**-2.11**	**.039**
Phase_Order	1.04	-.01 (.07)	-0.16	.871
Driving	1.30	.09 (.08)	1.14	.260
ASRS_Inattentive	3.17	-.01 (.08)	-0.08	.936
ASRS_Hyperactive	3.31	.05 (.09)	0.57	.573
Diagnosis	2.35	.01 (.08)	0.14	.885
COHS	2.12	-.04 (.07)	-0.64	.528
Condition	**64.79**	**.31 (.07)**	**4.68**	**<.001**
ASRS_Inattentive x Condition	16.51	-.02 (.08)	-0.29	.774
ASRS_Hyperactive x Condition	13.60	-.05 (.08)	-0.56	.575
Diagnosis x Condition	2.16	.10 (.07)	1.51	.134
COHS x Condition	57.37	.12 (.07)	1.74	.085

Model Comparisons

Model	***R^2***	***Log likel.***	***χ^2***	***χ^2 sig.***	***ΔR^2***
Model 1	.03	79.53			
Model 2	.03	79.87	0.70	.952	<.01
Model 3	**.13**	**90.64**	**21.53**	**<.001**	**.10**
Model 4	.15	93.73	6.19	.186	.03

Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Bottom layer of table, Model Comparisons, depicts the predictive strength of

each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface.

ADHD symptom severity and habit disruption

Similarly, we altered our pre-registered model to prevent multicollinearity, and performed a linear mixed model to examine the link between ADHD symptomology and habit disruption (see Table 4). The pre-registered analysis that violated assumptions of non-multicollinearity can be found in the Appendix, Supplemental Table 10. In our corrected model, we input Gender, Phase_Order, and Driving experience into step one, where none significantly predicted outcome-sensitivity (all *p*s > .142), model R^2= 0.01. In step two, we added ASRS_Inattentive, ASRS_Hyperactive, Diagnosis, and COHS into the model, and found that none of these regressors yielded main effects on outcome-sensitivity (all *p*s > .162), and they did not significantly improve the predictive strength of the model, $\chi^2(4) = 3.19$, $p = .526$, R^2= 0.03, $\Delta R^2 = 0.01$. We input Feedback as a regressor in step three, which contributed significantly to predicting outcome-sensitivity, $\beta_{Feedback} = -0.28$, $t(103) = -4.13$, $p < .001$, and rendered the model a significant predictor of ΔNoGo_Accuracy, $\chi^2(1) = 17.10$, $p <$ .001, R^2= 0.11, $\Delta R^2 = 0.08$. We performed post-hoc paired-samples t-tests to confirm the

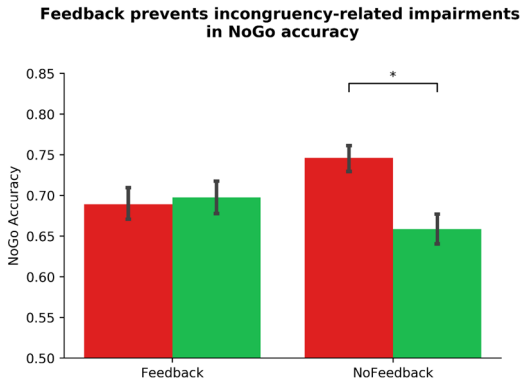


Figure 20. Dual monetary/performance feedback prevents the incongruency-related impairments in NoGo accuracy, breaking the habit. Participants exhibit no incongruency-related NoGo accuracy impairments after receiving cumulative performance and monetary feedback ($p = 616$). Without this feedback integration, participants exhibit a significant impairment in NoGo accuracy when the color-response mappings are incongruent with daily experiences ($p < .001$). The habit disruption effect of feedback is independent of ADHD symptom severity (see Table 4 for individual difference measure contributions to habit disruption). Color of bars reflects NoGo stimulus colors.

beneficial effect of dual feedback. We found that a significant NoGo accuracy impairment was evident in absence of dual feedback, $t(103) = 5.33$, $p < .001$, whereas the delivery of feedback yielded no significant accuracy impairments, $t(103) = -0.50$, $p = .616$ (see Figure 20). No individual difference measures' interaction regressor in step four significantly predicted outcome-sensitivity (all ps > .391, $\chi^2(4) = 1.56$, $p = .815$, $R^2 = 0.11$, $\Delta R^2 = 0.01$. These results suggest that the delivery of dual feedback indeed had a protective effect on outcome-sensitivity when managing familiar stimuli, albeit independent of ADHD symptom severity.

Table 4. Hierarchical Mixed Model of ADHD Symptomology and Habit Disruption: ΔNoGo_Accuracy.

Variable	***VIF***	***β***	***t***	***sig.***
Model 1				
Gender	1.01	.04 (.07)	0.60	.553
Phase_Order	1.01	.10 (.07)	1.48	.142
Driving	1.00	-.02 (.07)	-0.28	.779
Model 2				
Gender	1.08	.04 (.07)	0.62	.537
Phase_Order	1.04	.09 (.07)	1.25	.215
Driving	1.30	.02 (.08)	0.24	.807
ASRS_Inattentive	1.62	-.06 (.09)	-0.69	.491
ASRS_Hyperactive	1.71	.10 (.09)	1.12	.263
Diagnosis	1.30	-.05 (.08)	-0.61	.542
COHS	1.06	-.10 (.07)	-1.41	.162
Model 3				
Gender	1.08	.04 (.07)	0.64	.521
Phase_Order	1.04	.09 (.07)	1.30	.198
Driving	1.30	.02 (.08)	0.25	.799
ASRS_Inattentive	1.62	-.06 (.08)	-0.72	.474
ASRS_Hyperactive	1.71	.10 (.09)	1.17	.245
Diagnosis	1.30	-.05 (.08)	-0.64	.526
COHS	1.06	-.10 (.07)	-1.47	.146
Feedback	**1**	**-.28 (.07)**	**-4.13**	**<.001**
Model 4				
Gender	1.08	.04 (.07)	0.64	.525
Phase_Order	1.04	.09 (.07)	1.28	.202
Driving	1.30	.02 (.08)	0.25	.801
ASRS_Inattentive	3.15	-.06 (.08)	-0.71	.478
ASRS_Hyperactive	3.29	.10 (.09)	1.16	.250
Diagnosis	2.34	-.05 (.08)	-0.63	.530
COHS	2.10	-.10 (.07)	-1.45	.150
Feedback	**64.79**	**-.28 (.07)**	**-4.12**	**<.001**

ASRS_Inattentive x Feedback	16.49	.05 (.08)	0.62	.539
ASRS_Hyperactive x Feedback	13.58	-.01 (.08)	-0.16	.869
Diagnosis x Feedback	2.14	.02 (.07)	0.24	.811
COHS x Feedback	57.35	.06 (.07)	-0.86	.391

Model Comparisons

Model	R^2	*Log likel.*	χ^2	χ^2 *sig.*	ΔR^2
Model 1	.01	72.53			
Model 2	.03	74.13	3.19	.526	.01
Model 3	**.11**	**82.68**	**17.10**	**<.001**	**.08**
Model 4	.11	83.46	1.56	.815	.01

Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Bottom layer of table, Model Comparisons, depicts the predictive strength of each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface.

Supplementary analysis of ADHD symptom severity and well-learned habits

We performed identical analyses using ΔGo_Accuracy as DV and Subject as a random factor to capture the potential association between ADHD symptomology and a supplemental assay of outcome-sensitivity (see Table 5; see Appendix, Supplemental Table 11 for uncorrected model). Two participants' data were identified as outliers. Due to changes in statistical significance following outlier correction, we report our outlier-removed dataset below, highlighting any change in statistical significance due to outlier correction. Neither Gender, Phase_Order, or Driving experience predicted

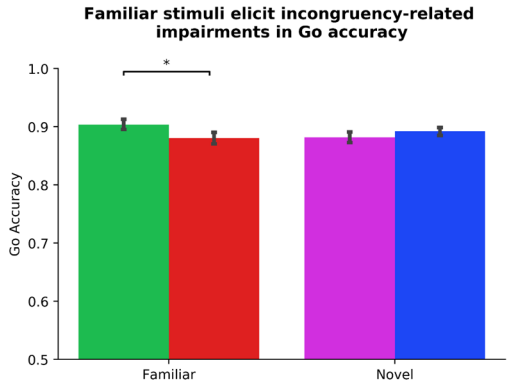


Figure 21. Familiar stimuli elicit incongruency-related impairments in Go accuracy. Analysis of our supplementary index of outcome-sensitivity, Go accuracy, yields evidence of habitual Go actions when managing familiar stimuli with color-response mappings that are incongruent with daily experiences ($p < .001$). In contrast, newly-learned Go/NoGo contingencies evoke no significant change in Go accuracy regardless of color-response mapping, indicating intact goal-directed performance ($p = .445$). The differential habit expression effect across Conditions depicted here is independent from ADHD symptom severity (see Table 5 for individual difference measure contributions to habit expression). Color of bars reflects Go stimulus colors.

ΔGo_Accuracy (all *p*s > .323), model R^2= 0.01. In step two, the Diagnosis regressor, which codes for the presence of a clinical ADHD diagnosis, made a significant contribution, $\beta_{Diagnosis}$ = 0.17, $t(94) = 2.11$, $p = .038$ (without outlier correction: $\beta_{Diagnosis}$ = 0.14, $t(96) = 1.80$, $p = .076$). Specifically, the presence of a diagnosis predicted more flexible Go actions. No other step two regressor significantly predicted ΔGo_Accuracy (all *p*s > .259) The step two model was not significantly improved from step one, $\chi^2(4) = 5.56$, $p = .235$, R^2= 0.04, $\Delta R^2 = 0.03$. The Condition regressor in step three served as a significant predictor, $\beta_{Condition}$ = 0.14, $t(101) = 2.07$, $p = .010$, improving the predictive strength of the model, $\chi^2(1) = 4.44$, $p = .035$, R^2= 0.06, $\Delta R^2 = 0.02$. Paired-samples t-tests revealed a significant Go accuracy impairment in the Familiar condition, $t(101) = 3.80$, $p < .001$, but not the Novel condition, $t(101) = -0.77$, $p = .445$ (see Figure 21). Lastly in step four, other than Diagnosis x Condition, $\beta_{Diagnosis \times Condition}$ = 0.19, $t(97) = 2.71$, $p = .008$, no individual difference measures significantly predicted ΔGo_Accuracy across the Familiar and Novel conditions (all other interaction *p*s > .125, $\chi^2(4) = 10.43$, $p = .034$, R^2= 0.10, $\Delta R^2 = 0.05$). Because we only had six individuals with an ADHD diagnosis, we refrain from further interpretations of the contribution of the Diagnosis regressor. These results suggest that Go accuracy is differentially affected by whether familiar or novel stimuli serve as Go/NoGo signals, and a significant impairment is evident when familiar contingencies are incongruent with daily experiences. However, the habitual Go actions elicited by our familiar stimuli are independent of ADHD symptom severity.

Table 5. Hierarchical Mixed Model of ADHD Symptomology and Habit Expression: ΔGo_Accuracy.

Variable	***VIF***	***β***	***t***	***sig.***
Model 1				
Gender	1.01	<.01 (.07)	-0.01	.997
Phase_Order	1.01	.07 (.07)	0.99	.323

Driving	1.00	.06 (.07)	0.88	.383
Model 2				
Gender	1.08	.02 (.07)	0.34	.731
Phase_Order	1.04	.08 (.07)	1.13	.260
Driving	1.30	-.01 (.08)	-0.23	.815
ASRS_Inattentive	1.59	.02 (.09)	0.22	.828
ASRS_Hyperactive	1.66	-.03 (.09)	-0.39	.699
Diagnosis	**1.30**	**.17 (.08)**	**2.11**	**.038**
COHS	1.06	-.03 (.07)	-0.41	.681
Model 3				
Gender	1.09	.02 (.07)	0.35	.729
Phase_Order	1.04	.08 (.07)	1.14	.256
Driving	1.30	-.02 (.08)	-0.24	.813
ASRS_Inattentive	1.59	.02 (.09)	0.22	.826
ASRS_Hyperactive	1.66	-.03 (.09)	-0.39	.696
Diagnosis	1.30	**.17 (.08)**	**2.13**	**.036**
COHS	1.06	-.03 (.07)	-0.42	.678
Condition	**1**	**.14 (.07)**	**2.07**	**.041**
Model 4				
Gender	1.09	.02 (.07)	0.35	.729
Phase_Order	1.04	.08 (.07)	1.16	.256
Driving	1.30	-.02 (.08)	-0.24	.813
ASRS_Inattentive	3.11	.02 (.09)	0.22	.826
ASRS_Hyperactive	3.23	-.03 (.09)	-0.40	.696
Diagnosis	2.35	**.17 (.08)**	**2.16**	**.036**
COHS	2.12	-.03 (.07)	-0.42	.678
Condition	**65.83**	**.14 (.07)**	**2.10**	**.038**
ASRS_Inattentive x Condition	16.65	-.08 (.08)	-0.95	.343
ASRS_Hyperactive x Condition	13.71	.04 (.09)	0.53	.599
Diagnosis x Condition	**2.16**	**.19 (.07)**	**2.71**	**.008**
COHS x Condition	57.08	.11 (.07)	1.55	.125

Model Comparisons

Model	***R²***	***Log likel.***	***χ²***	***χ² sig.***	***ΔR²***
Model 1	.01	218.44			
Model 2	.04	221.22	5.56	.235	.03
Model 3	**.06**	**223.44**	**4.44**	**.035**	**.02**
Model 4	**.10**	**228.65**	**10.40**	**.034**	**.05**

Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Model Comparisons layer depicts the predictive strength of each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface. Analyses have been outlier corrected, with resulting deviations highlighted in the text.

Supplementary analysis of ADHD symptom severity and habit disruption

We investigated habit disruption via mapping-related changes in Go accuracy using a similar mixed model (see Table 6). Our multicollinearity-corrected model identified two outliers (see Appendix, Supplemental Table 12 for uncorrected model). We report outlier-removed results below, accompanied by any changes in statistical

significance following outlier correction. In step one of the mixed model, no controlled regressors predicted ΔGo_Accuracy (all *p*s > .093), model R^2=.02. In step two, COHS was a near significant variable, β_{COHS} = -0.14, $t(94) = -1.95$, $p = .054$ (without outlier-correction: β_{COHS} = -0.08, $t(96) = -1.05$, $p = .296$), suggesting that a higher frequency of daily habits may predict more outcome-insensitive Go actions. Otherwise, no individual difference regressor served as a significant predictor of ΔGo_Accuracy (all other *p*s = .149), although the inclusion of step two regressors resulted in the Phase_Order variable to yield a near-significant p-value, $p = .066$. Step two regressors in aggregate yielded only a near-significant contribution on the DV, $\chi^2(4) = 8.56$, $p < .073$, $R^2 = 0.06$, $\Delta R^2 = 0.04$. In step three, the Feedback regressor significantly predicted outcome-sensitivity as indexed by ΔGo_Accuracy, $\beta_{Feedback}$ = -0.26, $t(101) = -4.07$, $p < .001$, improving the predictive strength of the model, $\chi^2(1) = 16.01$, $p < .001$, $R^2 = 0.13$, $\Delta R^2 = 0.07$. This finding suggests that outcome-sensitivity as assessed by ΔGo_Accuracy is differentially impacted depending on the availability of dual feedback. Indeed, a post-hoc paired-samples t-test confirms a significant impairment in Go accuracy when no feedback is delivered, $t(103)$

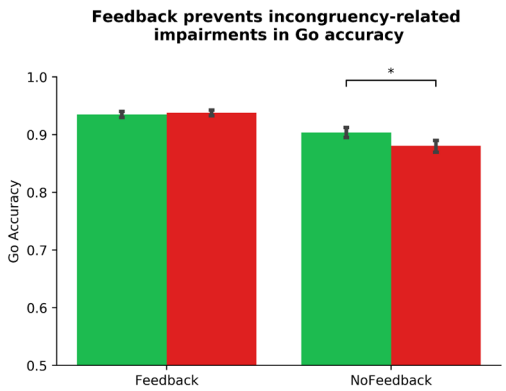


Figure 22. Dual monetary/performance feedback prevents the incongruency-related impairments in Go accuracy, breaking the habit. Similar to our NoGo accuracy results, analyses of the outcome-sensitivity measure of Go accuracy yield evidence for habit disruption due to cumulative performance and monetary feedback delivery. Participants exhibit no incongruency-related Go accuracy impairments after receiving dual feedback ($p = .573$). Without this feedback integration, participants exhibit a significant impairment in Go accuracy when the color-response mappings are incongruent with daily experiences ($p < .001$). The habit disruption effect of feedback is independent of ADHD symptom severity (see Table 6 for individual difference measure contributions to habit disruption). Color of bars reflects Go stimulus colors.

= 3.85, $p < .001$, whereas with feedback, no such impairment is evident, $t(103) = -0.56$, $p = .573$ (see Figure 22). In step four, we found that COHS x Feedback significantly predicted habit disruption, $\beta_{COHS \times Feedback} = -.16$, $t(97) = -2.46$, $p = .016$ (without outlier-correction: $p = .120$), suggesting that an increased daily habit frequency predicts a reduction in the beneficial effects of dual feedback in restoring goal-directed control. No other individual difference x Feedback regressor predicted habit disruption (all ps > .188, $\chi^2(4) = 9.70$, $p = .046$, $R^2 = 0.16$, $\Delta R^2 = 0.04$. Similar to our primary measure of outcome-sensitivity using NoGo accuracy, the protective effect of dual feedback on Go accuracy was independent from ADHD symptomology. However, we do observe a significant association between habitual tendencies in daily life and a difficulty in suppressing a well-learned habit.

Table 6. Hierarchical Mixed Model of ADHD Symptomology and Habit Disruption: ΔGo_Accuracy.

Variable	***VIF***	***β***	***t***	***sig.***
Model 1				
Gender	1.02	-.03 (.07)	-0.41	.684
Phase_Order	1.02	.12 (.07)	1.69	.093
Driving	1.00	-.03 (.07)	-0.49	.623
Model 2				
Gender	1.09	-.03 (.07)	-0.41	.679
Phase_Order	1.04	.13 (.07)	1.86	.066
Driving	1.30	-.05 (.08)	-0.63	.529
ASRS_Inattentive	1.59	.02 (.09)	0.27	.788
ASRS_Hyperactive	1.66	-.14 (.09)	-1.57	.121
Diagnosis	1.30	-.04 (.08)	-0.53	.598
COHS	1.06	-.14 (.07)	-1.95	.054
Model 3				
Gender	1.09	-.03 (.07)	-0.42	.678
Phase_Order	1.04	.13 (.07)	1.87	.065
Driving	1.30	-.05 (.08)	-0.63	.527
ASRS_Inattentive	1.59	.02 (.09)	0.27	.787
ASRS_Hyperactive	1.66	-.14 (.09)	-1.57	.119
Diagnosis	1.30	-.04 (.08)	-0.53	.596
COHS	1.06	-.14 (.07)	-1.96	.053
Feedback	**1**	**-.26 (.06)**	**-4.07**	**<.001**
Model 4				
Gender	1.09	-.03 (.07)	-0.41	.681

Phase_Order	1.04	.13 (.07)	1.85	.068
Driving	1.30	-.05 (.08)	-0.62	.531
ASRS_Inattentive	3.11	.02 (.09)	0.27	.789
ASRS_Hyperactive	3.23	-.14 (.09)	-1.56	.123
Diagnosis	2.35	-.04 (.08)	-0.53	.600
COHS	2.12	-.14 (.07)	-1.94	.056
Feedback	**65.83**	**-.26 (.06)**	**-4.22**	**<.001**
ASRS_Inattentive x Feedback	16.65	.10 (.08)	1.33	.188
ASRS_Hyperactive x Feedback	13.71	-.07 (.08)	-0.83	.410
Diagnosis x Feedback	2.16	<.01 (.06)	0.02	.984
COHS x Feedback	57.08	-.16 (.06)	-2.46	.016

Model Comparisons

Model	R^2	***Log likel.***	χ^2	χ^2 ***sig.***	ΔR^2
Model 1	.02	336.38			
Model 2	.06	340.66	8.56	.730	.04
Model 3	**.13**	**348.66**	**16.01**	**<.001**	**.07**
Model 4	**.16**	**353.52**	**9.70**	**.046**	**.04**

Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Model Comparisons layer depicts the predictive strength of each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface. Analyses have been outlier corrected, with resulting deviations highlighted in the text.

Exploratory analyses: Go RT and individual difference measures

We explored the potential association between prepotency to respond to the familiar Go stimulus and our individual difference measures of ADHD symptom severity (ASRS_Inattentive and ASRS_Hyperactive) and daily habit frequency (COHS). We reasoned that hyperactive individuals may exhibit a more pronounced prepotency to respond to Go stimuli, thus we were especially interested in the hyperactivity scale's association with RT. As hypothesized, we found a significant negative correlation between Go RT to the familiar green-Go color-response mapping and ASRS_Hyperactive, r = -.25, p = .030, Holm–Bonferroni corrected (Figure 23), suggesting that higher hyperactivity scores are associated with faster Go responses. This relationship between hyperactivity and response latency was not apparent when the Go signal was incongruent with lifelong experiences (red-Go r = -.05, p = 1, Holm–Bonferroni corrected), or when the Novel condition stimuli served as the Go signal (purple-Go r = -.12, p = .630; blue-Go r = -.10, p = .770, Holm–Bonferroni corrected). The association between familiar Go RT and ASRS_Hyperactive may suggest that individuals high in hyperactive symptoms may be exhibiting abnormally pronounced prepotency to stimuli that evoke habitual control.

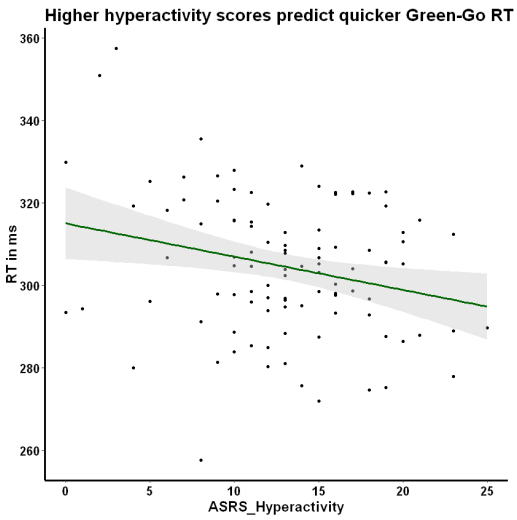


Figure 23. Hyperactivity symptom severity is negatively correlated with green-Go RT. Participants exhibit a significant negative correlation between hyperactivity symptoms and RT when responding to the familiar Go stimulus that is hypothesized to elicit prepotency. In other words, participants who score higher in hyperactivity make quicker Go responses when the contingencies are congruent with their daily representations. Pearson's r = -.25, p = .030, corrected for multiple comparisons using the Holm–Bonferroni method.

Discussion

The neurobehavioral evidence of atypical reward-related processes in ADHD, and the scarcity of strategies to restore potential behavioral rigidities, motivated us to examine the expression and disruption of well-learned habits as a function of ADHD symptom severity. To this end, we collected ADHD symptom severity metrics from a wide sample of participants in the general population and administered our Go/NoGo task that capitalizes on familiar green-Go/red-NoGo associations. Importantly, our incorporation of a motivational enhancement manipulation (i.e., cumulative performance and monetary feedback) permitted the study of habit expression and disruption. Our results replicate our recent documentation of familiar Go/NoGo stimuli evoking rigid habitual control, which is also rendered more flexible (i.e., goal-directed) with motivational enhancement (Ceceli et al., submitted). However, we found only modest support for the hypothesis of ADHD symptomology tracking behavioral rigidity and habit disruption. No measure of ADHD significantly predicted outcome-insensitivity as assayed by color-response mapping-related NoGo or Go accuracy impairments. Our exploratory analyses, however, supported our hypothesis of a significant association between pre-potency of habitual Go actions (i.e., familiar green-Go RT) and hyperactivity presentation. Furthermore, although not directly associated with ADHD, we also found a link between the frequency of habitual tendencies in daily life and habit disruption as indexed by our supplementary measure of outcome-sensitivity: mapping-related Go accuracy impairments. This significant association between daily habit frequency and difficulty breaking well-learned Go associations lends further credence to the idea that the familiar associations we capitalize on are indeed related to well-established, ecologically relevant habits.

In both scientific reports and diagnostic criteria, ADHD is characterized by pronounced deficits in inhibitory control (American Psychiatric Association, 2013; Wodka et al., 2007). When taken together with the reward-related irregularities, we posited that ADHD may also be associated with an impaired motivational control system favoring habits over goal-directed behaviors. Our results do not support this hypothesis with our primary analyses, which could be due to a few key factors.

First, our study recruited participants from the general population and obtained a normal distribution of ADHD-related symptom severity, such that most participants in our sample did not reach the clinical threshold for an ADHD diagnosis. This approach contextualizes any potential ADHD-related impairment in motivational processes to a wider audience, thus expanding the applicability of our research. Consequentially, we are unable to sufficiently represent those who are most debilitated by the symptoms in question: individuals who meet the clinical threshold for ADHD. Any potential ADHD-related effect may therefore be weakened by the large proportion of individuals who present symptoms below the clinical threshold at magnitudes that do not impair daily functioning. Indeed, a study that recruited adults from the general population to examine ADHD symptomology-related inhibitory control disparities found only a modest association between symptom severity and Go/NoGo task accuracy with 440 participants (Polner et al., 2015). A study with a larger sample size (n = 1156) obtained from the general population pinpointed Go/NoGo impairments due to high ADHD-like symptoms, though these effects were sensitive to variations in task structure (e.g., speed and reward structure) (Kuntsi et al., 2009). Taken together with our results, although the ADHD–Go/NoGo impairment association is well-documented in clinical presentations of ADHD,

symptom-based approaches may not be sensitive to such effects in the general population. Nonetheless, although there may be disorder-specific factors playing a role in behavioral flexibility that are undetected here, we had reasoned that sampling indiscriminately—that is, without diagnostic cutoffs—could expand the generalizability of potential symptom-related anomalies to the public.

An alternative explanation for the absence of a strong link between motivational control and ADHD symptomology is the notion that individuals with ADHD-like symptoms may also have compensatory mechanisms that promote adaptive behavioral output. For instance, despite the strong evidence of response inhibition deficits in ADHD, attention compensation supported by parietal brain activity has been documented, resulting in comparable Go/NoGo task performance (Dillo et al., 2010). Brain maturation is another candidate for behavioral similarities in ADHD and NT populations. ADHD is associated with a delayed maturation of the prefrontal cortex (Shaw et al., 2007), a region that is critical for error detection, reversal learning, and conflict monitoring. These processes are crucial for optimal Go/NoGo task performance (Garavan et al., 2002; Zhang et al., 2016), especially one involving changes to color-response mappings. Accordingly, adults with ADHD may produce signs of intact Go/NoGo performance due to the maturations in prefrontal regions, compensating for potential impairments that may have been evident with a less mature cortex (Carmona et al., 2012). Another potential compensatory mechanism may be driven by ADHD medications that act on the brain's dopaminergic systems. We did not ascertain whether our participants—with or without ADHD—were taking ADHD medication. Methylphenidate, for instance, has been reported to enhance executive function in individuals with ADHD, as well as in NTs

(Linssen et al., 2014; Moeller et al., 2014; Schweitzer et al., 2004). These beneficial effects of ADHD medication on executive function have also been shown to extend beyond methylphenidate (Hosenbocus and Chahal, 2012). Our sample of adults with varying degrees of ADHD-related symptoms may be recruiting similar compensatory mechanisms that aid in maintaining goal-directed control. Future research that captures developmental and pharmacological aspects of ADHD and goal-directed control may elucidate which of these mechanisms plays a critical role in adaptive motivational control.

We reasoned that because hyperactive ADHD presentation is associated with the number of impulsivity-related items endorsed on the ASRS (Kessler et al., 2005b), participants exhibiting high hyperactivity may execute quicker, impulsive Go actions. Our green-Go RT data supported our hypothesis, in that hyperactivity scores correlated with quicker responses to the well-learned habit eliciting stimulus. It should be noted that this finding was the result of an exploratory analysis. Nonetheless, our finding of a significant response latency and hyperactivity association bridges the fields of motivation and ADHD. Impulsivity, a core element of the hyperactive presentation of ADHD, is also associated with reflexive behaviors to cues and heightened variability in response latency (Kirkeby and Robinson, 2005). The heightened pre-potency to respond to habitual cues tracked by our hyperactivity scale may suggest an overlap in the motivational and inhibitory mechanisms underlying hyperactivity in ADHD, potentially explaining the lapses in behavioral output that result in higher RT and accuracy variability (Kirkeby and Robinson, 2005; Tamm et al., 2012). In other words, if hyperactivity predicts quicker responses to well-learned stimuli and high RT variability, this effect may be due to

motivational and motor processes that are activated depending on past experience with the cue at hand. Future research will be imperative in effectively dissociating the motivational, attentional, and inhibitory processes that underlie response latency variability in ADHD.

Conclusions

ADHD is a heterogenous psychiatric condition with debilitating consequences to behavior, neural processing, and well-being. In this study, we aimed to reveal the potential irregularities in managing well-learned habits by sampling symptom severity information from the general population. Although we did not find a strong association between motivational control deficits and ADHD-related symptoms, our data replicate a previous report of well-learned habit expression and disruption, and allude to a link between hyperactivity and pre-potency to respond to well-learned Go stimuli. Taken together with previous reports of compensatory mechanisms aiding in Go/NoGo task performance in ADHD, delay in cortical maturation in ADHD yielding differential inhibitory processes across children and adults, and our sample largely comprising subclinical ADHD presentations, a full understanding of the potential link between ADHD and motivational control may require a neurobehavioral and developmental approach.

Chapter 5: General Discussion and Implications

ADHD is a multi-faceted disorder with hallmark symptoms of inattentiveness and hyperactivity (American Psychiatric Association, 2013). The disorder is also known to exhibit aberrant response inhibition, reward processing, pronounced impulsivity, and deficits in working memory (Banca et al., 2015; Barkley, 1997; Castellanos & Tannock, 2002; Holmes et al., 2010; Kuntsi et al., 2001; Modesto-Lowe et al., 2013; Schachar et al., 1993, 1993; Suskauer, Simmonds, Fotedar, et al., 2008; Wilbertz et al., 2012). Considering that the dysfunctional reward circuitry observed in ADHD results in aberrant reward processing and maladaptive interactions with the environment (Castellanos and Tannock, 2002), I set out to reveal in this dissertation whether the neurobehavioral systems regulating the control of motivated behaviors were compromised.

Three studies aimed to reveal the processes underlying the formation, expression, and disruption of habits in ADHD. In the first study (Chapter 2), despite behavioral similarities, an atypical neural signature of motivational control in ADHD was evident, marked by diminished corticostriatal communication, and an early recruitment of the posterior putamen region that is associated with stimulus-sensitivity (McNamee et al., 2015; Tricomi et al., 2009). The second study (Chapter 3) focused on the expression and disruption of well-learned habits, and introduced a novel paradigm that capitalizes on well-established, ecologically valid representations of habitual behavior. This study also highlighted the beneficial effects of performance information and extrinsic rewards on motivational control by utilizing them to restore goal-directed control. Lastly, the third study (Chapter 4) applied a novel paradigm and feedback manipulation to a large sample from the general population to study well-learned habit expression and disruption as a

function of ADHD symptomology. This task elicited outcome-insensitive habits, and the feedback manipulation restored goal-directed control, albeit both independent of ADHD symptom severity. However, a modest link between habitual action execution prepotency (i.e., habitual Go action latency) and degree of hyperactivity was evident. In sum, these studies achieve three major goals: (1) highlighting the necessity of a motivational approach in empirical inquiries of ADHD by uncovering neural anomalies that may be precursors to behavioral rigidities, (2) providing the field the ability to capture a comprehensive snapshot of motivational control via well-learned habits, and (3) introducing motivational enhancement as a key strategy for restoring goal-directed control—a finding not only important for ADHD, but a variety of clinical disorders that exhibit reward-related impairments.

Habits in the context of ADHD

Throughout this dissertation, I tested using neural and behavioral tools whether individuals with ADHD had a proclivity to favor habits over goal-directed behaviors. Although the neural and behavioral manifestations are not in complete agreement with each other, an interesting "big picture" question remains: if ADHD should indeed present over-reliance on habitual control, what would this mean outside of the laboratory?

As reported in Chapters 2 and 4, ADHD presents an altered neural signature of top-down control, and hyperactivity may play a role in promoting habit execution. These findings allude to the notion that individuals with pronounced impulsivity may have difficulties controlling their goal-directed actions. This idea is supported by the prevalence of habit-related disorders that co-occur with ADHD, such as obsessive-compulsive and substance use (e.g., alcohol and stimulants) disorders (Downey et al.,

1997; McGough et al., 2005; Shekim et al., 1990). Like obsessive-compulsive disorder, addiction has been theorized to be a disorder of compulsivity (Everitt and Robbins, 2016; Griffiths et al., 2014). Specifically, Pavlovian and operant associations between drug cues, emotional and behavioral responses (e.g., craving and drug use), and the consequences of these S–R chains (e.g., euphoria) are strengthened to amplify the salience of triggering cues and contexts, eventually leading to compulsive drug use (Everitt and Robbins, 2005, 2016; Griffiths et al., 2014). While certain components of addiction may surely be goal-directed in nature (e.g., various nefarious acts committed by individuals in pursuit of acquiring drugs), research towards understanding the habitual aspects of addiction can inform translational endeavors, such as how to control cue-induced urges and maintain cessation goals. Indeed, even active smokers often report being unaware of their decision to smoke or having reached for a cigarette (Ikard et al., 1969). Those who are long-time quitters, as in individuals with cessation goals that conflict with a nicotine rush outcome, continue to experience cue-induced craving (Jager, 2003), supporting the notion that habit-like responses of craving and reaching for a cigarette upon cue prevail despite conflicting goals, or values associated with the outcome. In the context of ADHD, the development of compulsive, cue-sensitive processes may be pronounced, perhaps contributing to the prevalence of relevant comorbidities.

Potential contributors to ADHD's atypical neural signaling

If ADHD is marked by diminished corticostriatal signaling that can be a precursor to behavioral anomalies, what could be driving this irregularity? Chapter 2 found that compared to NTs, individuals with ADHD exhibited dampened signaling between the

dACC/mPFC and posterior putamen at cue onset following moderate S–R learning, as well as heightened posterior putamen recruitment. These regions have strong associations with value- and cue-based decision making and tracking reward value (Bush et al., 2002; Camille et al., 2011; O'Doherty, 2016; Smith et al., 2010; Zhang et al., 2016). A candidate mechanism underlying the corticostriatal abnormalities in the context of motivational control may be the utilization of cognitive resources, for which the prefrontal cortex is instrumental (Braver et al., 1997; Curtis and D'Esposito, 2003; Sawaguchi and Goldman-Rakic, 1991). Goal-directed control is known to be closely associated with working memory, in that individuals with higher cognitive resources in this domain are protected against factors that should otherwise render their behaviors habitual, such as stress (Otto et al., 2013; Quaedflieg et al., 2019). Furthermore, goal-directed control is incrementally recruited throughout development, possibly as a result of the expansion of cognitive resources and prefrontal maturation (Ceceli and Tricomi, 2018; Decker et al., 2016). In contrast, ADHD is associated with delayed prefrontal maturation, in that children with and without ADHD have differential slopes of neural development, potentially relating to delays in adaptive goal-directed control (Shaw et al., 2007). Further interrogations of habit formation and expression as a function of age can better inform these ideas and potentially bolster the cognitive resources hypothesis of motivational control in ADHD.

Neural correlates of well-learned habits and their disruption

Pioneering work in the late 20th century dissociating the habitual and goal-directed components of motivational control has spawned a vast literature of basic and applied studies of habits (Adams, 1982; Adams and Dickinson, 1981; Dickinson, 1985;

Dickinson and Balleine, 1994). The field has greatly benefited from the examinations of habit induction—indeed, the behavioral and neural shift from newly-acquired goal-directed behaviors to stimulus-triggered habits has been reliably mapped (Balleine and O'Doherty, 2009; Knowlton and Patterson, 2016; Yin and Knowlton, 2006). The prevalence of habit-based control has also been revealed in a variety of psychiatric disorders (Everitt et al., 2001; Griffiths et al., 2014). With that said, the studies in this dissertation were largely motivated by the disproportionate representation of habit formation versus the ecologically-relevant representations and disruptions of habits in the literature. For instance, habitual control has been demonstrated widely, even in horses and fruit flies (Brembs, 2011; Lansade et al., 2017), yet to the best of my knowledge, no study has examined the neural systems underlying the breaking of a habit. This dissertation has only scratched the surface of the processes underlying habit disruption. The neural correlates of habit disruption, for example, can have paramount implications for targeting effective treatment strategies for habit-based pathologies.

Candidate neural systems regulating the habit breaking process may overlap with regions involving cognitive control and goal-directed performance. Chapters 3 and 4 discussed well-learned habit disruption by reversing congruent color-response mappings to render them incongruent, such that participants needed to override their learned habit-like actions for optimal goal-directed performance. Possibly, the neural investigations of reversal learning may inform speculations about the habit breaking process. In a typical reversal learning paradigm, two cue–outcome (or cue–response–outcome, if the paradigm involves an instrumental response) contingencies are eventually reversed, and behavioral/physiological assays are used to detect reversal learning ability (Ghahremani

et al., 2010; Schiller et al., 2008; Zhang et al., 2016). Although the contingency change approach in Chapters 3 and 4 differs from reversal learning, in that participants undergoing the well-learned habit task are explicitly informed of the change in Go/NoGo contingencies, it is possible that these systems share neural circuitry. For instance, reversal learning in these studies have been associated with vmPFC, OFC and dACC activation—prefrontal regions that are also involved in the top-down control of motivated behaviors and cognitive control (Balleine and O'Doherty, 2009; Cole and Schneider, 2007; MacDonald et al., 2000).

An alternative explanation for the restoration of goal-directed behaviors may be the habit formation process occurring in reverse. If during habit formation, caudate and prefrontal cortex-regulated, outcome-driven processes shift towards putamen and motor cortex control and rely on triggering cues (Balleine and O'Doherty, 2009; Haber, 2003), perhaps restoring goal-directed control involves the reemergence of caudate and prefrontal cortex involvement during action control. Take, for example, the transcranial magnetic stimulation (TMS) of the lateral PFC and the its deleterious effects on model-based strategies and motor sequence learning (Desrochers et al., 2015; Smittenaar et al., 2013). Consider, also, the inactivation of the dorsomedial striatum (caudate) and the subsequent impairment in goal-directed control (Yin et al., 2005a), as well as dorsolateral striatum (putamen) lesions in rodents disrupting habit formation (Yin et al., 2004). Together, these studies suggest that perhaps increased involvement from the prefrontal cortex and caudate (or decreased involvement from the putamen) may foster goal-directed control. Future examinations of prefrontal function and amplification of prefrontal neuronal activity (possibly via transcranial direct current stimulation, or

TDCS) may reveal whether this region is essential in disrupting habits and restoring goal-directed control. This perspective is especially promising, considering that TDCS of the lateral PFC has been successfully used in enhancing working memory, and in a randomized clinical trial, alleviating cue-induced nicotine cravings (Fregni et al., 2005, 2008).

Lastly, these hypotheses involving the neural systems of habit disruption are also relevant for ADHD. Chapter 2 reports increased hippocampal activity following moderate S–R learning at cue onset in the ADHD group. Possibly, ADHD may be associated with declarative, hippocampal influences that aid in maintaining goal-directed performance, especially given the diminished corticostriatal connectivity. The lack of *a priori* declarative predictions limits such interpretations, although targeting the hippocampus to better understand ADHD's potential compensatory systems in future research may yield interesting findings. Such projects can also motivate similar examinations of behavioral flexibility in other disorders that exhibit corticostriatal irregularities (e.g., major depressive disorder and bipolar disorder; Satterthwaite et al., 2015).

Feedback as a tool to overcome habits

Chapters 3 and 4 introduce motivational enhancement as a powerful tool in restoring goal-directed control. Specifically, providing performance tracking information paired with a monetary incentive, possibly by amplifying the salience of a goal (e.g., task proficiency or maximizing monetary gain), induced behavioral flexibility. When further unpacked, Chapter 3, Experiment 2 reports that performance feedback alone is not successful in disrupting habits, and only modestly affects outcome-sensitivity. However,

evidence in Chapter 3, Experiment 3 suggests that the delivery of dual feedback is indeed successful. Performance information provided a baseline score of task performance for individuals, and possibly, only those who were intrinsically motivated (i.e., motivated by self-related reasons, such as task-competence and interest) benefited from the feedback. The significant improvement effect seen in Chapter 3, Experiment 3 may be the product of a global motivational boost, in that the delivery of dual feedback benefited both intrinsically and extrinsically motivated individuals. Specifically, dual feedback contained rewards that could be perceived as intrinsic (e.g., performance information that can be used towards tracking task competence) or extrinsic (e.g., money). However, it remains unknown whether monetary incentives alone could disrupt habits. This is a particularly interesting question for motivation researchers, as extrinsic rewards have been known to undermine intrinsic motivation, in that someone who is intrinsically motivated to perform a task may be negatively affected by receiving extrinsic rewards, diminishing self-generated motivation (Miller, 1988; Ryan and Deci, 2000). An alternative account posits that when reinforcement is indeed *rewarding*, contingent on task performance and not merely engaging in the task, and attainable, extrinsic rewards may not necessarily have a detrimental effect on intrinsic motivation (Dickinson, 1989). Examining the effects of monetary incentives in isolation, with measures that capture individual differences in motivation, can test whether either of these accounts provide more insight into the habit disruption process. Particularly, such an investigation can probe whether degree of habit disruption relies on type of motivation and reward.

Additionally, the reported effect of habit disruption may be relevant for remediating compulsive behaviors in the clinical setting. The motivational enhancement

manipulation via dual feedback shows significant overlap with feedback-based treatments of various disorders. For instance, contingency management (CM) is a therapeutic tool that capitalizes on similar motivational processes. During CM-based treatment, the patient is reinforced for performing desirable behaviors (i.e., substance use cessation, healthy eating, etc.), akin to models of operant conditioning (Prendergast et al., 2006). This method of behavioral therapy has yielded significant abstinence effects in the treatment of alcohol (Petry et al., 2000), cocaine (Epstein et al., 2003), and opioid abuse (Petry and Carroll, 2013; Petry and Martin, 2002). These beneficial effects also extend beyond abstinence therapy, as CM has also been documented to improve obesity treatment and the promotion of fitness goals (Volpp et al., 2008; Weinstock et al., 2008). Given that both CM and our motivational enhancement manipulations rely on inducing behavioral change via feedback, the approach in Chapters 3 and 4 may be useful in providing a mechanistic insight into these processes. Specifically, I asserted in these studies that dual feedback reactivates outcome representations in otherwise stimulus-dependent behaviors. Further work in this domain of habit disruption via motivational enhancement can supplement CM and other feedback-driven clinical practices by testing the underlying mechanisms of motivation in driving treatment efficacy (e.g., the neural correlates of habit disruption and its sustainability).

Thinking beyond disorder classifications

The three studies in this dissertation tackled habit formation, expression, and disruption in ADHD using two distinct approaches: the typical disorder classification method using the DSM (American Psychiatric Association, 2013), and a symptom-based method that captures behavioral manifestations in relation to symptom severity.

Chapter 2 assessed ADHD strictly per DSM guidelines and excluded participants with psychiatric comorbidities and non-ADHD-related medication regimens. Additionally, individuals with ADHD were demographically matched to their NT counterparts. Such a strict disorder classification approach serves several advantages. First, it ensured that the results can be attributed specifically to disparities in ADHD, minimizing the influence of other psychiatric illnesses. Second, this study attempted to capture the neurobehavioral systems of ADHD in absence of transient or enduring effects of non-ADHD related psychoactive medication use. Lastly, having closely matched ADHD and NT participants, behavioral and neural contrasts were performed with confidence, minimizing age, gender, and working memory-related confound effects.

Despite these advantages, participant selection based on stringent disorder classifications and exclusion criteria yields difficulties in research. For example, ADHD often manifests alongside other psychiatric disorders (McGough et al., 2005). It can be reasoned that among these excluded comorbidities, such as substance use and obsessive-compulsive disorder, there may be overlooked components that contribute to reward and motivation-related processes in ADHD. Importantly, with prevalence rates as high as 50% (Jensen and Steinhausen, 2015; McGough et al., 2005), a snapshot of ADHD that excludes comorbidities and their respective psychoactive treatment regimens may be unlike the real-world manifestation of the disorder. Such discrepancies between laboratory samples and the population at large may introduce two major issues. First, if the effect of interest (e.g., habit formation in ADHD) is related to the presence of a comorbid disorder, these effects may not be observable due to the stringent exclusion procedures. Second, the effects that stem from such recruitment protocols may weaken

the generalizability of the research, where the study sample does not sufficiently represent the clinical population (Sharp et al., 2016).

In concordance with the disadvantages of the typical disorder classification method outlined here, Chapter 4 approached the expression and disruption of habits in ADHD by borrowing from the National Institute of Mental Health's RDoC initiative. The core principles of RDoC are centered on measuring symptom dimensions, agnostic of DSM guidelines for any particular disorder (Insel et al., 2010). For instance, RDoC utilizes matrices of systems and levels of analysis. These matrices quantify domains such as positive valence (e.g., reward valuation, prediction error, etc.) via varying levels of analyses (e.g., behavior, molecular, genetic, etc.; Insel et al., 2010; Sharp et al., 2016). Chapter 4 presents a study in which a large sample from the general population was recruited, in absence of DSM's disorder categorization guidelines or any strict exclusion criteria. Continuous variables were employed to measure symptoms of inattention and hyperactivity, along with demographic information serving as controlled variables that can also contribute to explaining variability in behavior across the sample. Thus, this study was able to improve the generalizability of the findings, as any association reported in the results would be representative of the general population, and not a case of ADHD that occurs in isolation of other factors.

Limitations and future directions

Although the studies in this dissertation examined habits using multiple modalities (e.g., behavior, survey, brain activity) and perspectives (e.g., habit formation, expression, disruption), these efforts could be improved and extended in several ways.

In Chapter 2, individuals with ADHD exhibited dampened functional connectivity between the dACC and posterior putamen, and heightened posterior putamen activity as a function of S–R training. Although this study captured corticostriatal alterations in ADHD, these differences in brain function and connectivity did not accompany behavioral effects. Specifically, the ADHD group did not exhibit pronounced habit formation (i.e., the development of outcome-insensitivity). As mentioned in the "Thinking beyond disorder classifications" section, disorders of compulsivity that are often comorbid with ADHD may be driving potential habit effects. Because participants with other psychiatric diagnoses were excluded, the potential contribution of comorbidities on habit formation remains an open question. Furthermore, a developmental approach may also provide an explanation for the null behavioral findings. Due to the delayed cortical maturation seen in ADHD (Shaw et al., 2007), it may be the case that adults with ADHD (especially high functioning individuals in a predominantly collegiate sample) have neurally disparate, yet behaviorally intact motivational control. The developmental trajectory of motivational control in ADHD is an especially fascinating avenue of research. Goal-directed control has been documented to improve throughout development in the general population (Decker et al., 2016), and an aberrance in this process in ADHD would provide new insights into the lifespan of the disorder. Lastly, an alternative explanation for the behavioral similarities across ADHD and NT groups is that the induction of outcome devaluation via selective satiety of a food reward may not be the optimal approach to examine habits in humans. The devaluation paradigm relies on subjective reports of selective satiety, and may be susceptible to demand

characteristics. Although extensively employed in rodents, it might not be capturing the human habit experience with similar efficacy (Ceceli and Tricomi, 2018).

Chapter 3 and 4 utilized a novel paradigm that capitalized on existing associations to demonstrate and disrupt well-learned habits. I compared participants' behavioral flexibility to changes in Go and NoGo contingencies when managing familiar (green-Go, red-NoGo) or novel (purple-Go, blue-NoGo) color-response mappings. Because green and red should have strong associations with Go and NoGo behaviors, managing these signals when they are incongruent with daily experiences should produce outcome-insensitive, habit-like behaviors, in relation to the novel stimuli that presumably have no strong behavioral associations. Thus, the task relies on whether these visual stimuli elicit differences in action execution. Importantly, the stimuli are solely distinguished by color (see the Appendix, Supplemental Figure 1), in that a difficulty in identifying perceptually similar colors would complicate the interpretation of the findings. Chapter 3 deployed a between-subject design, in that two independent samples managed the red/green and blue/purple) task stimuli. Although this design would ensure that participants did not mistake red for any other color (since red would only be seen by the group undergoing the task with familiar stimuli), it does not account for potential similarities between purple and blue. To ensure that participants successfully distinguished these potentially similar colors (i.e., discerning red, blue, and purple), I administered a color discrimination task comprising the stimuli in the well-learned habit task (see the Appendix, Supplemental Figure 2). The details of this manipulation check can be found in the Appendix, Chapter 5 Supplement. In brief, participants were able to identify the

colors in question with similar accuracy, suggesting that the effects we report in Chapters 3 and 4 are not related to difficulties in identifying the task stimuli.

Conclusions

Neurobiological evidence suggests that ADHD is associated with reward-related deficits, as well as impairments in motivation that are tracked by assays of dopaminergic function in the brain (Castellanos and Tannock, 2002; Volkow et al., 2011). This dissertation examined whether ADHD was also associated with neurobehavioral abnormalities during the control of motivated behaviors. This approach contributes to the literature by highlighting corticostriatal alterations that may serve as ADHD endophenotypes. I also developed novel tools to demonstrate and disrupt habits that are suitable for examining not only ADHD, but a variety of disorders that exhibit reward-related symptoms. In sum, this work bolsters motivation as an important target for elucidating ADHD's pathophysiology, and may foster translational research to improve treatment strategies in clinical contexts—namely in remediating maladaptively rigid behaviors.

References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* 34, 77–98. doi:10.1080/14640748208400878.

Adams, C. D., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* 33, 109–121. doi:10.1080/14640748108400816.

Adriaanse, M. A., de Ridder, D. T. D., and de Wit, J. B. F. (2009). Finding the Critical Cue: Implementation Intentions to Change One's Diet Work Best When Tailored to Personally Relevant Reasons for Unhealthy Eating. *Pers. Soc. Psychol. Bull.* 35, 60–71. doi:10.1177/0146167208325612.

Adriaanse, M. A., Vinkers, C. D. W., de Ridder, D. T. D., Hox, J. J., and de Wit, J. B. F. (2011). Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. *Appetite* 56, 183–193. doi:10.1016/j.appet.2010.10.012.

Alvares, G. A., Balleine, B. W., and Guastella, A. J. (2014). Impairments in Goal-Directed Actions Predict Treatment Response to Cognitive-Behavioral Therapy in Social Anxiety Disorder. *PLOS ONE* 9, e94778. doi:10.1371/journal.pone.0094778.

Alvares, G. A., Balleine, B. W., Whittle, L., and Guastella, A. J. (2016). Reduced goal-directed action control in autism spectrum disorder. *Autism Res. Off. J. Int. Soc. Autism Res.* 9, 1285–1293. doi:10.1002/aur.1613.

Antrop, I., Stock, P., Verté, S., Wiersema, J. R., Baeyens, D., and Roeyers, H. (2006). ADHD and delay aversion: the influence of non-temporal stimulation on choice for delayed rewards. *J. Child Psychol. Psychiatry* 47, 1152–1158. doi:10.1111/j.1469-7610.2006.01619.x.

Balleine, B. W., and O'Doherty, J. P. (2009). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology* 35, 48–69. doi:10.1038/npp.2009.131.

Banca, P., Voon, V., Vestergaard, M. D., Philipiak, G., Almeida, I., Pocinho, F., et al. (2015). Imbalance in habitual versus goal directed neural systems during symptom provocation in obsessive-compulsive disorder. *Brain J. Neurol.* 138, 798–811. doi:10.1093/brain/awu379.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol. Bull.* 121, 65–94.

Barkley, R. A. (2005). *Attention-Deficit Hyperactivity Disorder, Third Edition: A Handbook for Diagnosis and Treatment*. Guilford Press.

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., et al. (2012). Reading Salt Activates Gustatory Brain Regions: fMRI Evidence for Semantic Grounding in a Novel Sensory Modality. *Cereb. Cortex N. Y. NY* 22, 2554–2563. doi:10.1093/cercor/bhr324.

Baum, W. M. (1993). Performances on ratio and interval schedules of reinforcement: Data and theory. *J. Exp. Anal. Behav.* 59, 245–264. doi:10.1901/jeab.1993.59-245.

Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *NeuroImage* 20, 1052–1063. doi:10.1016/S1053-8119(03)00435-X.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory. *NeuroImage* 5, 49–62. doi:10.1006/nimg.1996.0247.

Brembs, B. (2011). Spontaneous decisions and operant conditioning in fruit flies. *Behav. Processes* 87, 157–164. doi:10.1016/j.beproc.2011.02.005.

Bush, G., Frazier, J. A., Rauch, S. L., Seidman, L. J., Whalen, P. J., Jenike, M. A., et al. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop. *Biol. Psychiatry* 45, 1542–1552.

Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., et al. (2002). Dorsal anterior cingulate cortex: A role in reward-based decision making. *Proc. Natl. Acad. Sci. U. S. A.* 99, 523–528. doi:10.1073/pnas.012470999.

Camille, N., Tsuchida, A., and Fellows, L. K. (2011). Double Dissociation of Stimulus-Value and Action-Value Learning in Humans with Orbitofrontal or Anterior Cingulate Cortex Damage. *J. Neurosci.* 31, 15048–15052.

Carmona, S., Hoekzema, E., Ramos-Quiroga, J. A., Richarte, V., Canals, C., Bosch, R., et al. (2012). Response inhibition and reward anticipation in medication-naïve adults with attention-deficit/hyperactivity disorder: A within-subject case-control neuroimaging study. *Hum. Brain Mapp.* 33, 2350–2361. doi:10.1002/hbm.21368.

Carmona, S., Vilarroya, O., Bielsa, A., Trèmols, V., Soliva, J. C., Rovira, M., et al. (2005). Global and regional gray matter reductions in ADHD: a voxel-based morphometric study. *Neurosci. Lett.* 389, 88–93. doi:10.1016/j.neulet.2005.07.020.

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., et al. (2008). Cingulate-Precuneus Interactions: A New Locus of Dysfunction in Adult Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* 63, 332–337. doi:10.1016/j.biopsych.2007.06.025.

Castellanos, F. X., and Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nat. Rev. Neurosci.* 3, 617–628. doi:10.1038/nrn896.

Ceceli, A. O., and Tricomi, E. (2018). Habits and goals: a motivational perspective on action control. *Curr. Opin. Behav. Sci.* 20, 110–116. doi:10.1016/j.cobeha.2017.12.005.

Chun, M. M., and Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat. Neurosci.* 2, 844–847. doi:10.1038/12222.

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., and Petersen, S. E. (2014). Intrinsic and Task-Evoked Network Architectures of the Human Brain. *Neuron* 83, 238–251. doi:10.1016/j.neuron.2014.05.014.

Cole, M. W., and Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage* 37, 343–360. doi:10.1016/j.neuroimage.2007.03.071.

Costa Dias, T. G., Wilson, V. B., Bathula, D. R., Iyer, S. P., Mills, K. L., Thurlow, B. L., et al. (2013). Reward circuit connectivity relates to delay discounting in children with attention-deficit/hyperactivity disorder. *Eur. Neuropsychopharmacol.* 23, 33–45. doi:10.1016/j.euroneuro.2012.10.015.

Cubillo, A., Halari, R., Smith, A., Taylor, E., and Rubia, K. (2012). A review of fronto-striatal and fronto-cortical brain abnormalities in children and adults with Attention Deficit Hyperactivity Disorder (ADHD) and new evidence for dysfunction in adults with ADHD during motivation and attention. *Cortex J. Devoted Study Nerv. Syst. Behav.* 48, 194–215. doi:10.1016/j.cortex.2011.04.007.

Cunill, R., Castells, X., Tobias, A., and Capellà, D. (2015). Pharmacological treatment of attention deficit hyperactivity disorder with co-morbid drug dependence. *J. Psychopharmacol. (Oxf.)* 29, 15–23. doi:10.1177/0269881114544777.

Curtis, C. E., and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423. doi:10.1016/S1364-6613(03)00197-9.

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

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi:10.1038/nn1560.

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

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *J. Pers. Soc. Psychol.* 18, 105–115. doi:10.1037/h0030644.

Decker, J. H., Otto, A. R., Daw, N. D., and Hartley, C. A. (2016). From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. *Psychol. Sci.* 27, 848–858. doi:10.1177/0956797616639301.

Deichmann, R., Gottfried, J. A., Hutton, C., and Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage* 19, 430–441. doi:10.1016/S1053-8119(03)00073-9.

Delgado, M. R. (2007). Reward-Related Responses in the Human Striatum. *Ann. N. Y. Acad. Sci.* 1104, 70–88. doi:10.1196/annals.1390.002.

Delorme, C., Salvador, A., Valabrègue, R., Roze, E., Palminteri, S., Vidailhet, M., et al. (2016). Enhanced habit formation in Gilles de la Tourette syndrome. *Brain J. Neurol.* 139, 605–615. doi:10.1093/brain/awv307.

Desrochers, T. M., Chatham, C. H., and Badre, D. (2015). The necessity of rostrolateral prefrontal cortex for higher-level sequential behavior. *Neuron* 87, 1357–1368. doi:10.1016/j.neuron.2015.08.026.

Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. *R. Soc. Lond. Philos. Trans. Ser. B* 308, 67–78. doi:10.1098/rstb.1985.0010.

Dickinson, A. (1989). The detrimental effects of extrinsic reinforcement on "Intrinsic motivation." *Behav. Anal.* 12, 1–15.

Dickinson, A., and Balleine, B. (1994). Motivational control of goal-directed action. *Anim. Learn. Behav.* 22, 1–18. doi:10.3758/BF03199951.

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., and Boakes, R. A. (1995). Motivational control after extended instrumental training. *Anim. Learn. Behav.* 23, 197–206. doi:10.3758/BF03199935.

Dietrich, A., de Wit, S., and Horstmann, A. (2016). General Habit Propensity Relates to the Sensation Seeking Subdomain of Impulsivity But Not Obesity. *Front. Behav. Neurosci.* 10, 213. doi:10.3389/fnbeh.2016.00213.

Dillo, W., Göke, A., Prox-Vagedes, V., Szycik, G. R., Roy, M., Donnerstag, F., et al. (2010). Neuronal correlates of ADHD in adults with evidence for compensation

strategies – a functional MRI study with a Go/No-Go paradigm. *GMS Ger. Med. Sci.* 8. doi:10.3205/000098.

Dolan, R. J., and Dayan, P. (2013). Goals and Habits in the Brain. *Neuron* 80, 312–325. doi:10.1016/j.neuron.2013.09.007.

Douglas, V. I., and Parry, P. A. (1983). Effects of reward on delayed reaction time task performance of hyperactive children. *J. Abnorm. Child Psychol.* 11, 313–326.

Downey, K. K., Stelson, F. W., Pomerleau, O. F., and Giordani, B. (1997). Adult attention deficit hyperactivity disorder: psychological test profiles in a clinical population. *J. Nerv. Ment. Dis.* 185, 32–38.

Edel, M.-A., Enzi, B., Witthaus, H., Tegenthoff, M., Peters, S., Juckel, G., et al. (2013). Differential reward processing in subtypes of adult attention deficit hyperactivity disorder. *J. Psychiatr. Res.* 47, 350–356. doi:10.1016/j.jpsychires.2012.09.026.

Epstein, D. H., Hawkins, W. E., Covi, L., Umbricht, A., and Preston, K. L. (2003). Cognitive-behavioral therapy plus contingency management for cocaine use: Findings during treatment and across 12-month follow-up. *Psychol. Addict. Behav.* 17, 73–82. doi:10.1037/0893-164X.17.1.73.

Everitt, B. J., Dickinson, A., and Robbins, T. W. (2001). The neuropsychological basis of addictive behaviour. *Brain Res. Rev.* 36, 129–138. doi:10.1016/S0165-0173(01)00088-1.

Everitt, B. J., and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489. doi:10.1038/nn1579.

Everitt, B. J., and Robbins, T. W. (2016). Drug Addiction: Updating Actions to Habits to Compulsions Ten Years On. *Annu. Rev. Psychol.* 67, 23–50. doi:10.1146/annurev-psych-122414-033457.

Fayyad, J., De Graaf, R., Kessler, R. C., Alonso, J., Angermeyer, M., Demyttenaere, K., et al. (2007). Cross-national prevalence and correlates of adult attention-deficit hyperactivity disorder. *Br. J. Psychiatry J. Ment. Sci.* 190, 402–409. doi:10.1192/bjp.bp.106.034389.

Finn, P. R., Justus, A., Mazas, C., and Steinmetz, J. E. (1999). Working memory, executive processes and the effects of alcohol on Go/No-Go learning: testing a model of behavioral regulation and impulsivity. *Psychopharmacology (Berl.)* 146, 465–472.

Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., et al. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Exp. Brain Res.* 166, 23–30. doi:10.1007/s00221-005-2334-6.

Fregni, F., Liguori, P., Fecteau, S., Nitsche, M. A., Pascual-Leone, A., and Boggio, P. S. (2008). Cortical stimulation of the prefrontal cortex with transcranial direct current stimulation reduces cue-provoked smoking craving: A randomized, sham-controlled study. *J. Clin. Psychiatry* 69, 32–40. doi:10.4088/JCP.v69n0105.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6, 218–229. doi:10.1006/nimg.1997.0291.

Frodl, T., and Skokauskas, N. (2012). Meta-analysis of structural MRI studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects. *Acta Psychiatr. Scand.* 125, 114–126. doi:10.1111/j.1600-0447.2011.01786.x.

Furukawa, E., Bado, P., Tripp, G., Mattos, P., Wickens, J. R., Bramati, I. E., et al. (2014). Abnormal Striatal BOLD Responses to Reward Anticipation and Reward Delivery in ADHD. *PLOS ONE* 9, e89129. doi:10.1371/journal.pone.0089129.

Galvan, A., Hare, T. A., Davidson, M., Spicer, J., Glover, G., and Casey, B. J. (2005). The Role of Ventral Frontostriatal Circuitry in Reward-Based Learning in Humans. *J. Neurosci.* 25, 8650–8656. doi:10.1523/JNEUROSCI.2431-05.2005.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. a. P., and Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *NeuroImage* 17, 1820–1829.

Garner, D. M., Olmsted, M. P., Bohr, Y., and Garfinkel, P. E. (1982). The Eating Attitudes Test: psychometric features and clinical correlates. *Psychol. Med.* 12, 871–878. doi:10.1017/S0033291700049163.

Ghahremani, D. G., Monterosso, J., Jentsch, J. D., Bilder, R. M., and Poldrack, R. A. (2010). Neural Components Underlying Behavioral Flexibility in Human Reversal Learning. *Cereb. Cortex* 20, 1843–1852. doi:10.1093/cercor/bhp247.

Gillan, C. M., Apergis-Schoute, A. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Fineberg, N. A., et al. (2015). Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am. J. Psychiatry* 172, 284–293. doi:10.1176/appi.ajp.2014.14040525.

Gillan, C. M., Robbins, T. W., Sahakian, B. J., van den Heuvel, O. A., and van Wingen, G. (2016). The role of habit in compulsivity. *Eur. Neuropsychopharmacol.* 26, 828–840. doi:10.1016/j.euroneuro.2015.12.033.

Gollwitzer, P. M. (1999). Implementation Intentions : Strong Effects of Simple Plans. *Am. Psychol.* 54, 493–503. doi:10.1037/0003-066X.54.7.493.

Greene, A. J., Gross, W. L., Elsinger, C. L., and Rao, S. M. (2007). Hippocampal differentiation without recognition: An fMRI analysis of the contextual cueing task. *Learn. Mem. Cold Spring Harb. N* 14, 548–553. doi:10.1101/lm.609807.

Griffiths, K. R., Morris, R. W., and Balleine, B. W. (2014). Translational studies of goal-directed action as a framework for classifying deficits across psychiatric disorders. *Front. Syst. Neurosci.* 8, 101. doi:10.3389/fnsys.2014.00101.

Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* 26, 317–330. doi:10.1016/j.jchemneu.2003.10.003.

Harackiewlcz, J. M. (1979). The Effects of Reward Contingency and Performance Feedback on Intrinsic Motivation. *J. Pers. Soc. Psychol.* 37, 1352–63.

Hogarth, L., Chase, H. W., and Baess, K. (2012). Impaired goal-directed behavioural control in human impulsivity. *Q. J. Exp. Psychol. 2006* 65, 305–316. doi:10.1080/17470218.2010.518242.

Holland, R. W., Aarts, H., and Langendam, D. (2006). Breaking and creating habits on the working floor: A field-experiment on the power of implementation intentions. *J. Exp. Soc. Psychol.* 42, 776–783. doi:10.1016/j.jesp.2005.11.006.

Holmes, J., Gathercole, S. E., Place, M., Dunning, D. L., Hilton, K. A., and Elliott, J. G. (2010). Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Appl. Cogn. Psychol.* 24, 827–836. doi:10.1002/acp.1589.

Hosenbocus, S., and Chahal, R. (2012). A Review of Executive Function Deficits and Pharmacological Management in Children and Adolescents. *J. Can. Acad. Child Adolesc. Psychiatry* 21, 223–229.

Ikard, F. F., Green, D. E., and Horn, D. (1969). A Scale to Differentiate between Types of Smoking as Related to the Management of Affect. *Int. J. Addict.* 4, 649–659. doi:10.3109/10826086909062040.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 167, 748–751. doi:10.1176/appi.ajp.2010.09091379.

Jager, W. (2003). *Breaking 'bad habits': a dynamical perspective on habit formation and change*. L. Hendrickx, W. Jager, L. Steg (Eds.). University of Groningen, Groningen: Liber Amicorum for Charles Vlek Human Decision Making and Environmental Perception — Understanding and Assisting Human Decision Making in Real-life Settings.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17, 825–841.

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.

Jensen, C. M., and Steinhausen, H.-C. (2015). Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *Atten. Deficit Hyperact. Disord.* 7, 27–38. doi:10.1007/s12402-014-0142-1.

Johansen, E. B., Killeen, P. R., Russell, V. A., Tripp, G., Wickens, J. R., Tannock, R., et al. (2009). Origins of altered reinforcement effects in ADHD. *Behav. Brain Funct.* 5, 7. doi:10.1186/1744-9081-5-7.

Kessler, R. C., Adler, L. A., Barkley, R. A., Biederman, J., Conners, C. K., Faraone, S. V., et al. (2005a). Patterns and predictors of attention-deficit/hyperactivity disorder persistence into adulthood: results from the national comorbidity survey replication., Patterns and predictors of ADHD persistence into adulthood: Results from the National Comorbidity Survey Replication. *Biol. Psychiatry Biol. Psychiatry* 57, 57, 1442, 1442–1451. doi:10.1016/j.biopsych.2005.04.001, 10.1016/j.biopsych.2005.04.001.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., et al. (2005b). The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychol. Med.* 35, 245–256. doi:10.1017/S0033291704002892.

Kirkeby, B. S., and Robinson, M. D. (2005). Impulsive behavior and stimulus–response variability in choice reaction time. *J. Res. Personal.* 39, 263–277. doi:10.1016/j.jrp.2004.04.001.

Kleinsorge, T., and Rinkenauer, G. (2012). Effects of monetary incentives on task switching. *Exp. Psychol.* 59, 216–226. doi:10.1027/1618-3169/a000146.

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119, 254–284. doi:10.1037/0033-2909.119.2.254.

Knowlton, B. J., and Patterson, T. K. (2016). Habit Formation and the Striatum. *Curr. Top. Behav. Neurosci.* doi:10.1007/7854_2016_451.

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., and Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fmri. *Neuroreport* 12, 3683–3687.

Kolar, D., Keller, A., Golfinopoulos, M., Cumyn, L., Syer, C., and Hechtman, L. (2008). Treatment of adults with attention-deficit/hyperactivity disorder. *Neuropsychiatr. Dis. Treat.* 4, 389–403.

Kuntsi, J., Oosterlaan, J., and Stevenson, J. (2001). Psychological mechanisms in hyperactivity: I. Response inhibition deficit, working memory impairment, delay aversion, or something else? *J. Child Psychol. Psychiatry* 42, 199–210.

Kuntsi, J., Wood, A. C., Meere, J. V. D., and Asherson, P. (2009). Why cognitive performance in ADHD may not reveal true potential: Findings from a large population-based sample. *J. Int. Neuropsychol. Soc.* 15, 570–579. doi:10.1017/S135561770909081X.

Lansade, L., Marchand, A. R., Coutureau, E., Ballé, C., Polli, F., and Calandreau, L. (2017). Personality and predisposition to form habit behaviours during instrumental conditioning in horses (Equus caballus). *PloS One* 12, e0171010. doi:10.1371/journal.pone.0171010.

Lei, D., Du, M., Wu, M., Chen, T., Huang, X., Du, X., et al. (2015). Functional MRI reveals different response inhibition between adults and children with ADHD. *Neuropsychology* 29, 874–881. doi:10.1037/neu0000200.

Linnebank, F. E., Kindt, M., and de Wit, S. (2018). Investigating the balance between goal-directed and habitual control in experimental and real-life settings. *Learn. Behav.* 46, 306–319. doi:10.3758/s13420-018-0313-6.

Linssen, A. M. W., Sambeth, A., Vuurman, E. F. P. M., and Riedel, W. J. (2014). Cognitive effects of methylphenidate in healthy volunteers: a review of single dose studies. *Int. J. Neuropsychopharmacol.* 17, 961–977. doi:10.1017/S1461145713001594.

Luman, M., Oosterlaan, J., Knol, D. L., and Sergeant, J. A. (2008). Decision-making in ADHD: sensitive to frequency but blind to the magnitude of penalty? *J. Child Psychol. Psychiatry* 49, 712–722. doi:10.1111/j.1469-7610.2008.01910.x.

Luman Marjolein, Oosterlaan Jaap, Knol Dirk L., and Sergeant Joseph A. (2008). Decision-making in ADHD: sensitive to frequency but blind to the magnitude of penalty? *J. Child Psychol. Psychiatry* 49, 712–722. doi:10.1111/j.1469-7610.2008.01910.x.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., and Carter, C. S. (2000). Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science* 288, 1835–1838. doi:10.1126/science.288.5472.1835.

Makris, N., Biederman, J., Valera, E. M., Bush, G., Kaiser, J., Kennedy, D. N., et al. (2007). Cortical Thinning of the Attention and Executive Function Networks in Adults with Attention-Deficit/Hyperactivity Disorder. *Cereb. Cortex* 17, 1364–1375. doi:10.1093/cercor/bhl047.

Marx, I., Höpcke, C., Berger, C., Wandschneider, R., and Herpertz, S. C. (2013). The Impact of Financial Reward Contingencies on Cognitive Function Profiles in Adult ADHD. *PLOS ONE* 8, e67002. doi:10.1371/journal.pone.0067002.

Marx, I., Hübner, T., Herpertz, S. C., Berger, C., Reuter, E., Kircher, T., et al. (2010). Cross-sectional evaluation of cognitive functioning in children, adolescents and

young adults with ADHD. *J. Neural Transm.* 117, 403–419. doi:10.1007/s00702-009-0345-3.

McGough, J. J., Smalley, S. L., McCracken, J. T., Yang, M., Del'Homme, M., Lynn, D. E., et al. (2005). Psychiatric comorbidity in adult attention deficit hyperactivity disorder: findings from multiplex families. *Am. J. Psychiatry* 162, 1621–1627. doi:10.1176/appi.ajp.162.9.1621.

McKim, T. H., Bauer, D. J., and Boettiger, C. A. (2016). Addiction History Associates with the Propensity to Form Habits. *J. Cogn. Neurosci.* 28, 1024–1038. doi:10.1162/jocn_a_00953.

McNamee, D., Liljeholm, M., Zika, O., and O'Doherty, J. P. (2015). Characterizing the Associative Content of Brain Structures Involved in Habitual and Goal-Directed Actions in Humans: A Multivariate fMRI Study. *J. Neurosci.* 35, 3764–3771. doi:10.1523/JNEUROSCI.4677-14.2015.

Miller, K. A. (1988). Review of Intrinsic Motivation and Self-Determination in Human Behavior. *Contemp. Sociol.* 17, 253–253. doi:10.2307/2070638.

Modesto-Lowe, V., Chaplin, M., Soovajian, V., and Meyer, A. (2013). Are motivation deficits underestimated in patients with ADHD? A review of the literature. *Postgrad. Med.* 125, 47–52. doi:10.3810/pgm.2013.07.2677.

Moeller, S. J., Honorio, J., Tomasi, D., Parvaz, M. A., Woicik, P. A., Volkow, N. D., et al. (2014). Methylphenidate Enhances Executive Function and Optimizes Prefrontal Function in Both Health and Cocaine Addiction. *Cereb. Cortex* 24, 643–653. doi:10.1093/cercor/bhs345.

Montague, W. E., and Webber, C. E. (1965). Effects of Knowledge of Results and Differential Monetary Reward on Six Uninterrupted Hours of Monitoring. *Hum. Factors* 7, 173–180. doi:10.1177/001872086500700209.

Morris, L. S., Baek, K., and Voon, V. (2017). Distinct cortico-striatal connections with subthalamic nucleus underlie facets of compulsivity. *Cortex J. Devoted Study Nerv. Syst. Behav.* 88, 143–150. doi:10.1016/j.cortex.2016.12.018.

Morris, R. W., Quail, S., Griffiths, K. R., Green, M. J., and Balleine, B. W. (2015). Corticostriatal Control of Goal-Directed Action Is Impaired in Schizophrenia. *Biol. Psychiatry* 77, 187–195. doi:10.1016/j.biopsych.2014.06.005.

Mostofsky, S. H., Schafer, J. G. B., Abrams, M. T., Goldberg, M. C., Flower, A. A., Boyce, A., et al. (2003). fMRI evidence that the neural basis of response inhibition is task-dependent. *Cogn. Brain Res.* 17, 419–430. doi:10.1016/S0926-6410(03)00144-7.

Naito, E., and Matsumura, M. (1996). Movement-related potentials associated with motor inhibition under different preparatory states during performance of two visual stop

signal paradigms in humans. *Neuropsychologia* 34, 565–573. doi:10.1016/0028-3932(95)00140-9.

Naqvi, N. H., and Bechara, A. (2010). The insula and drug addiction: an interoceptive view of pleasure, urges, and decision-making. *Brain Struct. Funct.* 214, 435–450. doi:10.1007/s00429-010-0268-7.

Naqvi, N. H., Rudrauf, D., Damasio, H., and Bechara, A. (2007). Damage to the Insula Disrupts Addiction to Cigarette Smoking. *Science* 315, 531–534. doi:10.1126/science.1135926.

Natsheh, J. Y., and Shiflett, M. W. (2015). The Effects of Methylphenidate on Goal-directed Behavior in a Rat Model of ADHD. *Front. Behav. Neurosci.* 9, 326. doi:10.3389/fnbeh.2015.00326.

Natsheh, J. Y., and Shiflett, M. W. (2018). Dopaminergic Modulation of Goal-Directed Behavior in a Rodent Model of Attention-Deficit/Hyperactivity Disorder. *Front. Integr. Neurosci.* 12, 45. doi:10.3389/fnint.2018.00045.

Norman, L. J., Carlisi, C., Lukito, S., Hart, H., Mataix-Cols, D., Radua, J., et al. (2016). Structural and Functional Brain Abnormalities in Attention-Deficit/Hyperactivity Disorder and Obsessive-Compulsive Disorder: A Comparative Meta-analysis. *JAMA Psychiatry* 73, 815–825. doi:10.1001/jamapsychiatry.2016.0700.

O'Doherty, J. P. (2016). Multiple Systems for the Motivational Control of Behavior and Associated Neural Substrates in Humans. *Curr. Top. Behav. Neurosci.* 27, 291–312. doi:10.1007/7854_2015_386.

O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454. doi:10.1126/science.1094285.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., and Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Soc. Cogn. Affect. Neurosci.* 7, 604–609. doi:10.1093/scan/nss055.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., and Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci.* 110, 20941–20946. doi:10.1073/pnas.1312011110.

Otto, A. R., Skatova, A., Madlon-Kay, S., and Daw, N. D. (2015). Cognitive control predicts use of model-based reinforcement learning. *J. Cogn. Neurosci.* 27, 319–333. doi:10.1162/jocn_a_00709.

Patton, J. H., Stanford, M. S., and Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *J. Clin. Psychol.* 51, 768–774. doi:10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1.

Petry, N. M., and Carroll, K. M. (2013). Contingency Management Is Efficacious in Opioid-Dependent Outpatients Not Maintained on Agonist Pharmacotherapy. *Psychol. Addict. Behav. J. Soc. Psychol. Addict. Behav.* 27, 1036–1043. doi:10.1037/a0032175.

Petry, N. M., and Martin, B. (2002). Low-cost contingency management for treating cocaine- and opioid-abusing methadone patients. *J. Consult. Clin. Psychol.* 70, 398–405. doi:10.1037//0022-006X.70.2.398.

Petry, N. M., Martin, B., Cooney, J. L., and Kranzler, H. R. (2000). Give them prizes, and they will come: contingency management for treatment of alcohol dependence. *J. Consult. Clin. Psychol.* 68, 250–257.

Plichta, M. M., and Scheres, A. (2014). Ventral–striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: A meta-analytic review of the fMRI literature. *Neurosci. Biobehav. Rev.* 38, 125–134. doi:10.1016/j.neubiorev.2013.07.012.

Plichta, M. M., Vasic, N., Wolf, R. C., Lesch, K.-P., Brummer, D., Jacob, C., et al. (2009). Neural Hyporesponsiveness and Hyperresponsiveness During Immediate and Delayed Reward Processing in Adult Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* 65, 7–14. doi:10.1016/j.biopsych.2008.07.008.

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., et al. (2005). The Neural Correlates of Motor Skill Automaticity. *J. Neurosci.* 25, 5356–5364. doi:10.1523/JNEUROSCI.3880-04.2005.

Polli, F. E., Barton, J. J. S., Cain, M. S., Thakkar, K. N., Rauch, S. L., and Manoach, D. S. (2005). Rostral and dorsal anterior cingulate cortex make dissociable contributions during antisaccade error commission. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15700–15705. doi:10.1073/pnas.0503657102.

Polner, B., Aichert, D., Macare, C., Costa, A., and Ettinger, U. (2015). Gently restless: association of ADHD-like traits with response inhibition and interference control. *Eur. Arch. Psychiatry Clin. Neurosci.* 265, 689–699. doi:10.1007/s00406-014-0531-7.

Prendergast, M., Podus, D., Finney, J., Greenwell, L., and Roll, J. (2006). Contingency management for treatment of substance use disorders: a meta-analysis. *Addict. Abingdon Engl.* 101, 1546–1560. doi:10.1111/j.1360-0443.2006.01581.x.

Qiu, A., Crocetti, D., Adler, M., Mahone, E. M., Denckla, M. B., Miller, M. I., et al. (2009). Basal Ganglia Volume and Shape in Children With Attention Deficit Hyperactivity Disorder. *Am. J. Psychiatry* 166, 74–82. doi:10.1176/appi.ajp.2008.08030426.

Quaedflieg, C. W. E. M., Stoffregen, H., Sebalo, I., and Smeets, T. (2019). Stress-induced impairment in goal-directed instrumental behaviour is moderated by

baseline working memory. *Neurobiol. Learn. Mem.* doi:10.1016/j.nlm.2019.01.010.

Radenbach, C., Reiter, A. M. F., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.-J., et al. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology* 53, 268–280. doi:10.1016/j.psyneuen.2014.12.017.

Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A., and Todd, R. D. (2010). Sex and age differences in Attention-Deficit/Hyperactivity Disorder symptoms and diagnoses: Implications for DSM-V and ICD-11. *J. Am. Acad. Child Adolesc. Psychiatry* 49, 217-28.e1–3.

Reber, J., Feinstein, J. S., O'Doherty, J. P., Liljeholm, M., Adolphs, R., and Tranel, D. (2017). Selective impairment of goal-directed decision-making following lesions to the human ventromedial prefrontal cortex. *Brain J. Neurol.* 140, 1743–1756. doi:10.1093/brain/awx105.

Reiter, A. M. F., Deserno, L., Kallert, T., Heinze, H.-J., Heinz, A., and Schlagenhauf, F. (2016). Behavioral and Neural Signatures of Reduced Updating of Alternative Options in Alcohol-Dependent Patients during Flexible Decision-Making. *J. Neurosci. Off. J. Soc. Neurosci.* 36, 10935–10948. doi:10.1523/JNEUROSCI.4322-15.2016.

Rickham, P. P. (1964). HUMAN EXPERIMENTATION. CODE OF ETHICS OF THE WORLD MEDICAL ASSOCIATION. DECLARATION OF HELSINKI. *Br. Med. J.* 2, 177.

Rosch, K. S., Mostofsky, S. H., and Nebel, M. B. (2018). ADHD-related sex differences in fronto-subcortical intrinsic functional connectivity and associations with delay discounting. *J. Neurodev. Disord.* 10, 34. doi:10.1186/s11689-018-9254-9.

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp. Educ. Psychol.* 25, 54–67. doi:10.1006/ceps.1999.1020.

Satterthwaite, T. D., Kable, J. W., Vandekar, L., Katchmar, N., Bassett, D. S., Baldassano, C. F., et al. (2015). Common and Dissociable Dysfunction of the Reward System in Bipolar and Unipolar Depression. *Neuropsychopharmacology* 40, 2258–2268. doi:10.1038/npp.2015.75.

Sawaguchi, T., and Goldman-Rakic, P. S. (1991). D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251, 947–950. doi:10.1126/science.1825731.

Schachar, R. J., Tannock, R., and Logan, G. (1993). Inhibitory control, impulsiveness, and attention deficit hyperactivity disorder. *Clin. Psychol. Rev.* 13, 721–739. doi:10.1016/S0272-7358(05)80003-0.

Scheres, A., Milham, M. P., Knutson, B., and Castellanos, F. X. (2007). Ventral Striatal Hyporesponsiveness During Reward Anticipation in Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* 61, 720–724. doi:10.1016/j.biopsych.2006.04.042.

Schiller, D., Levy, I., Niv, Y., LeDoux, J. E., and Phelps, E. A. (2008). From Fear to Safety and Back: Reversal of Fear in the Human Brain. *J. Neurosci.* 28, 11517–11525. doi:10.1523/JNEUROSCI.2265-08.2008.

Schneider, M. F., Krick, C. M., Retz, W., Hengesch, G., Retz-Junginger, P., Reith, W., et al. (2010). Impairment of fronto-striatal and parietal cerebral networks correlates with attention deficit hyperactivity disorder (ADHD) psychopathology in adults — A functional magnetic resonance imaging (fMRI) study. *Psychiatry Res. Neuroimaging* 183, 75–84. doi:10.1016/j.pscychresns.2010.04.005.

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Curr. Opin. Neurobiol.* 7, 191–197. doi:10.1016/S0959-4388(97)80007-4.

Schultz, W., Tremblay, L., and Hollerman, J. R. (2000). Reward Processing in Primate Orbitofrontal Cortex and Basal Ganglia. *Cereb. Cortex* 10, 272–283. doi:10.1093/cercor/10.3.272.

Schwabe, L., Dalm, S., Schächinger, H., and Oitzl, M. S. (2008). Chronic stress modulates the use of spatial and stimulus-response learning strategies in mice and man. *Neurobiol. Learn. Mem.* 90, 495–503. doi:10.1016/j.nlm.2008.07.015.

Schwabe, L., Tegenthoff, M., Höffken, O., and Wolf, O. T. (2012). Simultaneous Glucocorticoid and Noradrenergic Activity Disrupts the Neural Basis of Goal-Directed Action in the Human Brain. *J. Neurosci.* 32, 10146–10155. doi:10.1523/JNEUROSCI.1304-12.2012.

Schwabe, L., and Wolf, O. T. (2009). Stress prompts habit behavior in humans. *J. Neurosci. Off. J. Soc. Neurosci.* 29, 7191–7198. doi:10.1523/JNEUROSCI.0979-09.2009.

Schwabe, L., and Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology* 35, 977–986. doi:10.1016/j.psyneuen.2009.12.010.

Schwabe, L., and Wolf, O. T. (2011). Stress increases behavioral resistance to extinction. *Psychoneuroendocrinology* 36, 1287–1293. doi:10.1016/j.psyneuen.2011.02.002.

Schweitzer, J. B., Lee, D. O., Hanford, R. B., Zink, C. F., Ely, T. D., Tagamets, M. A., et al. (2004). Effect of methylphenidate on executive functioning in adults with attention-deficit/hyperactivity disorder: normalization of behavior but not related brain activity. *Biol. Psychiatry* 56, 597–606. doi:10.1016/j.biopsych.2004.07.011.

Seidman, L. J., Biederman, J., Liang, L., Valera, E. M., Monuteaux, M. C., Brown, A., et al. (2011). Gray matter alterations in adults with attention-deficit/hyperactivity disorder identified by voxel based morphometry. *Biol. Psychiatry* 69, 857–866. doi:10.1016/j.biopsych.2010.09.053.

Seidman, L. J., Valera, E. M., Makris, N., Monuteaux, M. C., Boriel, D. L., Kelkar, K., et al. (2006). Dorsolateral Prefrontal and Anterior Cingulate Cortex Volumetric Abnormalities in Adults with Attention-Deficit/Hyperactivity Disorder Identified by Magnetic Resonance Imaging. *Biol. Psychiatry* 60, 1071–1080. doi:10.1016/j.biopsych.2006.04.031.

Sescousse, G., Redouté, J., and Dreher, J.-C. (2010). The architecture of reward value coding in the human orbitofrontal cortex. *J. Neurosci. Off. J. Soc. Neurosci.* 30, 13095–13104. doi:10.1523/JNEUROSCI.3501-10.2010.

Sethi, A., Voon, V., Critchley, H. D., Cercignani, M., and Harrison, N. A. (2018). A neurocomputational account of reward and novelty processing and effects of psychostimulants in attention deficit hyperactivity disorder. *Brain* 141, 1545–1557. doi:10.1093/brain/awy048.

Sharp, C., Fowler, J. C., Salas, R., Nielsen, D., Allen, J., Oldham, J., et al. (2016). Operationalizing NIMH Research Domain Criteria (RDoC) in naturalistic clinical settings. *Bull. Menninger Clin.* 80, 187–212. doi:10.1521/bumc.2016.80.3.187.

Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., et al. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 19649–19654. doi:10.1073/pnas.0707741104.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 Suppl 20, 22-33;quiz 34-57.

Shekim, W. O., Asarnow, R. F., Hess, E., Zaucha, K., and Wheeler, N. (1990). A clinical and demographic profile of a sample of adults with attention deficit hyperactivity disorder, residual state. *Compr. Psychiatry* 31, 416–425.

Shen, Y. J., and Chun, M. M. (2011). Increases in rewards promote flexible behavior. *Atten. Percept. Psychophys.* 73, 938–952. doi:10.3758/s13414-010-0065-7.

Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T. W., Beekman, A. T. F., Penninx, B. W. J. H., et al. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* 3, e337. doi:10.1038/tp.2013.107.

Sjoerds, Z., Dietrich, A., Deserno, L., de Wit, S., Villringer, A., Heinze, H.-J., et al. (2016). Slips of Action and Sequential Decisions: A Cross-Validation Study of

Tasks Assessing Habitual and Goal-Directed Action Control. *Front. Behav. Neurosci.* 10, 234. doi:10.3389/fnbeh.2016.00234.

Slusarek, M., Velling, S., Bunk, D., and Eggers, C. (2001). Motivational Effects on Inhibitory Control in Children With ADHD. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 355–363. doi:10.1097/00004583-200103000-00016.

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., and Huettel, S. A. (2010). Distinct Value Signals in Anterior and Posterior Ventromedial Prefrontal Cortex. *J. Neurosci. Off. J. Soc. Neurosci.* 30, 2490–2495. doi:10.1523/JNEUROSCI.3319-09.2010.

Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., and Dolan, R. J. (2013). Disruption of Dorsolateral Prefrontal Cortex Decreases Model-Based in Favor of Model-free Control in Humans. *Neuron* 80, 914. doi:10.1016/j.neuron.2013.08.009.

Snorrason, I., Lee, H. J., de Wit, S., and Woods, D. W. (2016). Are nonclinical obsessive-compulsive symptoms associated with bias toward habits? *Psychiatry Res.* 241, 221–223. doi:10.1016/j.psychres.2016.04.067.

Soares, J. M., Sampaio, A., Ferreira, L. M., Santos, N. C., Marques, F., Palha, J. A., et al. (2012). Stress-induced changes in human decision-making are reversible. *Transl. Psychiatry* 2, e131. doi:10.1038/tp.2012.59.

Sonuga-Barke, E. J., Taylor, E., Sembi, S., and Smith, J. (1992). Hyperactivity and delay aversion--I. The effect of delay on choice. *J. Child Psychol. Psychiatry* 33, 387–398.

Steinhausen, H.-C., Nøvik, T. S., Baldursson, G., Curatolo, P., Lorenzo, M. J., Rodrigues Pereira, R., et al. (2006). Co-existing psychiatric problems in ADHD in the ADORE cohort. *Eur. Child Adolesc. Psychiatry* 15 Suppl 1, I25-29. doi:10.1007/s00787-006-1004-y.

Stoy, M., Schlagenhauf, F., Schlochtermeier, L., Wrase, J., Knutson, B., Lehmkuhl, U., et al. (2011). Reward processing in male adults with childhood ADHD—a comparison between drug-naïve and methylphenidate-treated subjects. *Psychopharmacology (Berl.)* 215, 467–481. doi:10.1007/s00213-011-2166-y.

Ströhle, A., Stoy, M., Wrase, J., Schwarzer, S., Schlagenhauf, F., Huss, M., et al. (2008). Reward anticipation and outcomes in adult males with attention-deficit/hyperactivity disorder. *NeuroImage* 39, 966–972. doi:10.1016/j.neuroimage.2007.09.044.

Suskauer, S. J., Simmonds, D. J., Fotedar, S., Blankner, J. G., Pekar, J. J., Denckla, M. B., et al. (2008). Functional Magnetic Resonance Imaging Evidence for Abnormalities in Response Selection in Attention Deficit Hyperactivity Disorder:

Differences in Activation Associated with Response Inhibition but Not Habitual Motor Response. *J. Cogn. Neurosci.* 20, 478–493. doi:10.1162/jocn.2008.20032.

Tabachnick, B. G., and Fidell, L. S. (2007). *Using multivariate statistics, 5th ed*. Boston, MA: Allyn & Bacon/Pearson Education.

Tamm, L., Narad, M. E., Antonini, T. N., O'Brien, K. M., Hawk, L. W., and Epstein, J. N. (2012). Reaction Time Variability in ADHD: A Review. *Neurotherapeutics* 9, 500–508. doi:10.1007/s13311-012-0138-5.

Taylor, S. B., Anglin, J. M., Paode, P. R., Riggert, A. G., Olive, M. F., and Conrad, C. D. (2014). Chronic stress may facilitate the recruitment of habit- and addiction-related neurocircuitries through neuronal restructuring of the striatum. *Neuroscience* 280, 231–242. doi:10.1016/j.neuroscience.2014.09.029.

Tomasi, D., and Volkow, N. D. (2012). Abnormal Functional Connectivity in Children with Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* 71, 443–450. doi:10.1016/j.biopsych.2011.11.003.

Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi:10.1111/j.1460-9568.2009.06796.x.

Tricomi, E., Delgado, M. R., and Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron* 41, 281–292.

Umemoto, A., and Holroyd, C. B. (2015). Task-specific effects of reward on task switching. *Psychol. Res.* 79, 698–707. doi:10.1007/s00426-014-0595-z.

Urcelay, G. P., and Dalley, J. W. (2012). Linking ADHD, impulsivity, and drug abuse: a neuropsychological perspective. *Curr. Top. Behav. Neurosci.* 9, 173–197. doi:10.1007/7854_2011_119.

Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the Neural Substrates of Goal-Directed Learning in the Human Brain. *J. Neurosci.* 27, 4019–4026. doi:10.1523/JNEUROSCI.0564-07.2007.

Veldhuizen, M. G., Bender, G., Constable, R. T., and Small, D. M. (2007). Trying to Detect Taste in a Tasteless Solution: Modulation of Early Gustatory Cortex by Attention to Taste. *Chem. Senses* 32, 569–581. doi:10.1093/chemse/bjm025.

Verbruggen, F., and Logan, G. D. (2009). Automaticity of cognitive control: goal priming in response-inhibition paradigms. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1381–1388. doi:10.1037/a0016645.

Verhoeven, A. A. C., Kindt, M., Zomer, C. L., and de Wit, S. (2017). An experimental investigation of breaking learnt habits with verbal implementation intentions. *Acta Psychol. (Amst.).* doi:10.1016/j.actpsy.2017.05.008.

Volkow, N. D., Wang, G.-J., Kollins, S. H., Wigal, T. L., Newcorn, J. H., Telang, F., et al. (2009). Evaluating Dopamine Reward Pathway in ADHD. *JAMA J. Am. Med. Assoc.* 302, 1084–1091. doi:10.1001/jama.2009.1308.

Volkow, N. D., Wang, G.-J., Newcorn, J. H., Kollins, S. H., Wigal, T. L., Telang, F., et al. (2011). Motivation deficit in ADHD is associated with dysfunction of the dopamine reward pathway. *Mol. Psychiatry* 16, 1147–1154. doi:10.1038/mp.2010.97.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., and Loewenstein, G. (2008). Financial incentive-based approaches for weight loss: a randomized trial. *JAMA* 300, 2631–2637. doi:10.1001/jama.2008.804.

von Rhein, D., Beckmann, C. F., Franke, B., Oosterlaan, J., Heslenfeld, D. J., Hoekstra, P. J., et al. (2017). Network-level assessment of reward-related activation in patients with ADHD and healthy individuals. *Hum. Brain Mapp.* 38, 2359–2369. doi:10.1002/hbm.23522.

von Rhein, D., Cools, R., Zwiers, M. P., van der Schaaf, M., Franke, B., Luman, M., et al. (2015). Increased Neural Responses to Reward in Adolescents and Young Adults With Attention-Deficit/Hyperactivity Disorder and Their Unaffected Siblings. *J. Am. Acad. Child Adolesc. Psychiatry* 54, 394–402. doi:10.1016/j.jaac.2015.02.012.

Watson, P., and de Wit, S. (2018). Current limits of experimental research into habits and future directions - ScienceDirect. *Curr. Opin. Behav. Sci.* 20, 33–39.

Webb, T. L., Sheeran, P., and Luszczynska, A. (2009). Planning to break unwanted habits: Habit strength moderates implementation intention effects on behaviour change. *Br. J. Soc. Psychol.* 48, 507–523. doi:10.1348/014466608X370591.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale: WAIS-IV ; Technical and Interpretive Manual*. Pearson.

Weinstock, J., Barry, D., and Petry, N. M. (2008). Exercise-related activities are associated with positive outcome in contingency management treatment for substance use disorders. *Addict. Behav.* 33, 1072–1075. doi:10.1016/j.addbeh.2008.03.011.

Wilbertz, G., Tebartz van Elst, L., Delgado, M. R., Maier, S., Feige, B., Philipsen, A., et al. (2012). Orbitofrontal reward sensitivity and impulsivity in adult attention deficit hyperactivity disorder. *NeuroImage* 60, 353–361. doi:10.1016/j.neuroimage.2011.12.011.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., and Pennington, B. F. (2005). Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Biol. Psychiatry* 57, 1336–1346. doi:10.1016/j.biopsych.2005.02.006.

Wodka, E. L., Mahone, E. M., Blankner, J. G., Larson, J. C. G., Fotedar, S., Denckla, M. B., et al. (2007). Evidence that response inhibition is a primary deficit in ADHD. *J. Clin. Exp. Neuropsychol.* 29, 345–356. doi:10.1080/13803390600678046.

Wodushek, T. R., and Neumann, C. S. (2003). Inhibitory capacity in adults with symptoms of Attention Deficit/Hyperactivity Disorder (ADHD). *Arch. Clin. Neuropsychol.* 18, 317–330. doi:10.1016/S0887-6177(02)00152-X.

Wood, W., and Rünger, D. (2016). Psychology of Habit. *Annu. Rev. Psychol.* 67, 289–314. doi:10.1146/annurev-psych-122414-033417.

Worbe, Y., Palminteri, S., Savulich, G., Daw, N. D., Fernandez-Egea, E., Robbins, T. W., et al. (2016). Valence-dependent influence of serotonin depletion on model-based choice strategy. *Mol. Psychiatry* 21, 624–629. doi:10.1038/mp.2015.46.

Worbe, Y., Savulich, G., de Wit, S., Fernandez-Egea, E., and Robbins, T. W. (2015). Tryptophan Depletion Promotes Habitual over Goal-Directed Control of Appetitive Responding in Humans. *Int. J. Neuropsychopharmacol.* 18, pyv013. doi:10.1093/ijnp/pyv013.

Yee, D. M., Krug, M. K., Allen, A. Z., and Braver, T. S. (2016). Humans Integrate Monetary and Liquid Incentives to Motivate Cognitive Task Performance. *Front. Psychol.* 6. doi:10.3389/fpsyg.2015.02037.

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi:10.1038/nrn1919.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2005a). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.* 22, 505–512. doi:10.1111/j.1460-9568.2005.04219.x.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166, 189–196. doi:10.1016/j.bbr.2005.07.012.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., and Balleine, B. W. (2005b). The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513–523. doi:10.1111/j.1460-9568.2005.04218.x.

Zedelius, C. M., Veling, H., Bijleveld, E., and Aarts, H. (2012). Promising High Monetary Rewards for Future Task Performance Increases Intermediate Task Performance. *PLOS ONE* 7, e42547. doi:10.1371/journal.pone.0042547.

Zhang, Z., Mendelsohn, A., Manson, K. F., Schiller, D., and Levy, I. (2016). Dissociating Value Representation and Inhibition of Inappropriate Affective Response during Reversal Learning in the Ventromedial Prefrontal Cortex. *eNeuro* 2. doi:10.1523/ENEURO.0072-15.2015.

Appendix

Chapter 2 Supplement

Supplemental Table 1. Activation clusters and local maxima within contrasts (Task > Rest onset, Late > Early phase).

				MNI coordinates (mm)		
Cluster index	**Brain region labels (Hemisphere)**	**Cluster size in voxels**	**Z-stat**	**x**	**y**	**z**
PPI analysis, NT > ADHD						
1	Paracingulate gyrus (R)	804	3.65	8	50	16
	Anterior cingulate gyrus (R)		3.57	6	44	10
	Anterior cingulate gyrus (L)		3.47	-10	40	10
	Superior frontal gyrus (L)		3.15	-2	52	24
	Paracingulate gyrus (L)		3.09	-2	46	28
	Frontal pole (R)		2.75	2	66	16
Whole-brain analysis, ADHD > NT						
1	Hippocampus (L)	838	4.02	-12	-14	-20
	Posterior parahippocampal gyrus (L)		4.00	-14	-34	-20
	Anterior parahippocampal gyrus (L)		3.76	-12	-16	-24
	Posterior temporal fusiform cortex (L)		3.64	-30	-38	-18
	Brain stem (R)		3.60	6	-26	-34
	Cerebellum (L)		3.42	-8	-38	-16
	Posterior putamen (L)		3.09	-32	-26	0
2	Central opercular cortex (L)	749	4.01	-46	-12	18
	Insular cortex (L)		3.50	-38	2	-8
	Precentral gyrus (L)		3.48	-62	2	8
	Postcentral gyrus (L)		3.19	-58	-14	18
	Temporal pole (L)		3.10	-38	4	-26
	Planum polare (L)		3.02	-46	-8	0
	Parahippocampal gyrus (L)		2.80	-30	-4	-28

Local maxima within each cluster with Z-stat values greater than the cluster-defining threshold of 2.58 ($p < .005$, corrected to $p < .05$). When analyses yield multiple local maxima that belong to the same anatomical region, the coordinates depicted correspond to the activation with the higher Z-stat value. The Harvard-Oxford and Montreal Neurological Institute (MNI) Structural Atlases were used to map coordinates to their corresponding brain region labels. PPI: psychophysiological interaction, ADHD: attention-deficit/hyperactivity disorder, NT: neurotypical, L: left hemisphere (negative coordinates on the x-plane), R: right hemisphere (positive coordinates on the x-plane). Voxel dimensions: isotropic 3 mm.

Chapter 3 Supplement

<u>Experiment 1 – Omnibus regression to confirm mapping-related accuracy impairment</u>

Primary measure of outcome-sensitivity: NoGo accuracy

We derived a ΔNoGo_Accuracy (i.e., change in NoGo accuracy scores across mappings) DV to quantify the mapping-related impairment for each subject. This ΔNoGo_Accuracy variable serves as the primary measure of outcome-sensitivity, in that a greater impairment represents greater outcome-insensitivity. Specifically, difficulty overriding the Familiar Red–NoGo association for the Green–NoGo association indicates a cue-driven habit. In contrast, we would not expect a pronounced ΔNoGo_Accuracy value when participants manage Novel NoGo contingencies (i.e., blue–NoGo and purple–NoGo should yield similar accuracy scores). Participants with DV standardized residual values below -3.3 and above +3.3 were identified as outliers (Tabachnick & Fidell, 2007). In such cases, we performed identical analyses without outlier participants to verify robustness of findings, but only report these excluded analyses if outliers produced substantial changes in statistical significance.

We employed a hierarchical multiple regression model to extract the predictive strength of the between-group Condition variable while controlling for age, gender, order of mapping phase (i.e., whether a subject completed a particular color-response mapping first), and impulsivity. We entered the controlled Age, Gender, Order, and Impulsivity regressors into the first, and the Condition regressor of interest into the second step of the model. Therefore, our hierarchical multiple regression model yielded an R^2 change value (ΔR^2) for Condition, determining whether mapping-related impairments are predicted specifically by Condition (i.e., contingency change in Familiar versus Novel stimuli). We also derived the corresponding F_{change} value, which compares the predictive strength of the variables in the second step of the model with those in the first step (i.e., confirming whether ΔR^2 reflects a significant change in the model's predictive strength).

The regression model met the assumptions of normality and homoscedasticity. Multicollinearity tests produced negligible Variance Inflation Factors (VIF), confirming linearity assumptions of the regression (VIF for all variables < 1.09). Model 1, a linear combination of the controlled variables of Age, Gender, Order, and Impulsivity, did not significantly predict outcome-sensitivity: $F(4,45) = 0.46$, $p = .767$, and only explained 4% of the variance in ΔNoGo_Accuracy ($R^2 = .04$). Additionally, no controlled regressor independently predicted a change in ΔNoGo_Accuracy (all β coefficient *p*'s. > .05). In the second step of the regression, the inclusion of Condition as a regressor explained an additional 15.5% of the variance in outcome-sensitivity: $\beta_{Condition} = -.40$, $\Delta R^2 = .15$, $F_{change}(1,44) = 8.47$, $p = .006$—a significant contribution. Thus, the addition of the significant Condition regressor rendered the entirety of Model 2 a near-significant predictor of outcome-sensitivity, despite the null contributions from Age, Gender, Order, and Impulsivity: $F(5,44) = 2.12$, $p = .081$ (see Supplementary Table 2).

Supplemental Table 2. Summary of the Hierarchical Multiple Regression Model for Outcome-Insensitivity as Assayed by ΔNoGo_Accuracy.

Variable	***Toler.***	***VIF***	***B***	***SE***	***β***	***t***	***sig.***
Model 1							
Age	.98	1.02	0.57	0.82	.10	0.69	.493
Gender	.97	1.04	-2.69	5.08	-.08	-0.53	.599
Impulsivity	.93	1.07	-0.25	0.35	-.11	-0.71	.483
Order	.96	1.04	3.12	4.88	.95	0.64	.527
Model 2							
Age	.97	1.03	0.34	0.77	.06	0.44	.662
Gender	.96	1.05	-1.30	4.73	-.04	-0.27	.784
Impulsivity	.93	1.08	-0.33	0.33	-.14	-0.99	.326
Order	.96	1.04	2.88	4.52	.09	0.64	.527
Condition	.98	1.02	-13.06	4.50	-.40**	-2.91	**.006**

Model Summary Statistics						
Model	R^2	*F*	*F sig.*	ΔR^2	F_{change}	F_{change} *sig.*
Model 1	.04	0.46	.767			
Model 2	.19	2.12	.081	.15	8.47	**.006**

Note: Top layer of table depicts all regressors included in the hierarchical model and their respective statistics. Bottom layer of table, Model Summary Statistics, depicts the predictive strength of each model.

Delta R^2 (ΔR^2) and corresponding F_{change} values denote the specific improvement of Model 2 over Model 1 in predicting the dependent variable. Toler. = Tolerance; VIF = Variance Inflation Factor. Significant p-values (alpha = .05) depicted in bold typeface.

These results also suggest that the differential mapping-related impairment observed across Familiar and Novel conditions is not due to the order in which participants managed color-response mappings. The Order variable did not significantly predict ΔNoGo_Accuracy in our model ($\beta = .09$, $p = .527$). We found no interaction between factors of Order and Mapping in NoGo accuracy as a result of the repeated measures ANOVA: $F(1,48) = 0.35$, $p = .555$, $\eta_p^2 < .01$. We performed the same ANOVA separately in Familiar and Novel conditions and observed no significant interactions in either group (p's > .05).

Secondary measure of outcome-sensitivity: Go accuracy

We performed an identical omnibus regression using ΔGo_Accuracy as DV—the secondary assay of outcome-sensitivity. The regression model met the assumptions of normality and homoscedasticity. Multicollinearity tests produced negligible Variance Inflation Factors (VIF), confirming the assumption of linearity in the regression model (VIF for all variables < 1.09).

Collectively, the linear combination of Age, Gender, Order, and Impulsivity did not significantly predict mapping-related impairments in Go accuracy: $F(4,45) = 2.18$, $p = .087$. As depicted in Supplementary Table 3, closer examination of the individual regressors revealed a significant role played by Age, such that older participants suffered a greater mapping-related Go accuracy impairment: $\beta_{Age} = -.31$, $p = .027$. The inclusion of the Condition regressor significantly improved the predictive strength of the model in step 2: $\beta_{Condition} = -.27$, $\Delta R^2 = .07$, $F_{change}(1,44) = 4.10$, $p = .049$, and the significant

contribution of Age remained: β_{Age} = -.34, p = .014. Although age was a significant predictor of change in Go accuracy, because we had no *a priori* hypothesis, and the correlational direction of this relationship varied across conditions (Familiar Condition Pearson's r = .43; Novel Condition Pearson's r = -.70), we refrain from further age-related speculation.

Supplemental Table 3. Summary of the Hierarchical Multiple Regression Model for Outcome-Insensitivity as Assayed by ΔGo_Accuracy.

Variable	***Toler.***	***VIF***	***B***	***SE***	***β***	***t***	***sig.***
Model 1							
Age	.98	1.02	-0.60	0.26	-.31	-2.29*	**.027**
Gender	.97	1.04	1.83	1.61	.16	1.14	.261
Impulsivity	.93	1.07	-0.02	0.11	-.03	-0.20	.843
Order	.96	1.04	-2.02	1.55	-.18	-1.30	.199
Model 2							
Age	.97	1.03	-0.65	0.25	-.34*	-2.56	**.014**
Gender	.96	1.05	2.15	1.56	.19	1.38	.176
Impulsivity	.93	1.08	-0.04	0.11	-.05	-0.37	.716
Order	.96	1.04	-2.07	1.50	-.19	-1.38	.174
Condition	.98	1.02	-3.01	1.49	-.27**	-2.03	**.049**

Model Summary Statistics						
Model	R^2	F	F *sig.*	ΔR^2	F_{change}	F_{change} *sig.*
Model 1	.16	2.18	.087			
Model 2	.23	2.68	**.034**	.07	4.10	**.049**

Note: Top layer of table depicts all regressors included in the hierarchical model and their respective statistics. Bottom layer of table, Model Summary Statistics, depicts the predictive strength of each model. Delta R^2 (ΔR^2) and corresponding F_{change} values denote the specific improvement of Model 2 over Model 1 in predicting the dependent variable. Toler. = Tolerance; VIF = Variance Inflation Factor. Significant p-values (alpha = .05) depicted in bold typeface.

Similar to our primary assay of outcome-sensitivity, change in Go accuracy was not due to the order in which participants managed color-response mappings (β_{Order} = -.19, p = .174). There was no significant Order x Mapping interaction in Go accuracy: $F(1,48) = 2.26$, p = .140, η_p^2 = .04. We performed the same ANOVA separately in Familiar and Novel conditions and observed no significant interactions in either group (p's > .05).

Experiment 2 – Omnibus regression to illustrate mapping-related impairment while testing the effect of cumulative performance feedback

Primary measure of outcome-sensitivity: NoGo accuracy

Similar to Experiment 1, we derived a ΔNoGo_Accuracy DV to quantify the mapping-related impairment for each participant as the primary assay of outcome-sensitivity. We employed a hierarchical multiple regression model to extract the predictive strengths of the between-group regressors, Condition and Feedback, while controlling for age, gender, and impulsivity. The resulting ΔR^2 value for the contributions of Condition and Feedback determined whether mapping-related impairments are predicted specifically by Condition (i.e., Familiar versus Novel stimuli), and whether the cumulative performance feedback manipulation plays a role in affecting motivational control. A corresponding F_{change} value was derived to confirm whether ΔR^2 reflects a significant change in the model's predictive strength. Participants with DV standardized residual values below -3.3 and above 3.3 were identified as outliers (Tabachnick & Fidell, 2007). In such cases, we performed identical analyses without outlier participants to verify robustness of findings, but only report these excluded analyses if outliers produced substantial changes in statistical significance.

The regression model met the assumptions of normality and homoscedasticity. Multicollinearity tests produced negligible Variance Inflation Factors (VIF), confirming the assumption of linearity in the regression model (VIF for all variables < 1.04; see Supplementary Table 4).

Model 1, a linear combination of the controlled variables Age, Gender, and Impulsivity, did not significantly predict outcome-sensitivity: $F(3,96) = 0.18$, $p = .91$, and

only explained 0.6% of the variance in ΔNoGo_Accuracy. Additionally, no controlled regressor independently predicted a change in ΔNoGo_Accuracy (all β coefficient *p*'s. > .05; see Supplementary Table 4). In the second step of the regression, the inclusion of the Condition and Feedback regressors explained an additional 14.5% of the variance in outcome-sensitivity: $\beta_{Condition}$ = -.34, p = .001, $\beta_{Feedback}$ = .18, p = .07, ΔR^2 = .14, F_{change} (2,94) = 8.03, p = .001—rendering the entirety of Model 2 a significant predictor of outcome-sensitivity: $F(5,94) = 3.34$, p = .008 (see Supplementary Table 4).

Supplemental Table 4. Summary of the Hierarchical Multiple Regression Model for Outcome-Insensitivity as Assayed by ΔNoGo_Accuracy.

Variable	***Toler.***	***VIF***	***B***	***SE***	***β***	***t***	***sig.***
Model 1							
Age	.99	1.01	-0.31	0.58	-.05	-0.54	.590
Gender	.98	1.02	-0.31	3.75	-.01	-0.08	.935
Impulsivity	.98	1.02	0.13	0.27	.05	0.48	.633
Model 2							
Age	.98	1.02	-0.54	0.54	-.09	-1.00	.319
Gender	.96	1.04	1.10	3.53	.03	0.31	.756
Impulsivity	.98	1.02	0.17	0.25	.06	0.68	.497
Condition	.98	1.02	-11.74	3.29	-.34**	-3.57	**.001**
Feedback	.99	1.01	6.06	3.28	.18	1.85	.067

Model Summary Statistics						
Model	R^2	*F*	*F sig.*	ΔR^2	F_{change}	F_{change} *sig.*
Model 1	.01	0.18	.910			
Model 2	.15	3.34	**.008**	.14	8.03	**.001**

Note: Top layer of table depicts all regressors included in the hierarchical model and their respective statistics. Bottom layer of table, Model Summary Statistics, depicts the predictive strength of each model. Delta R^2 (ΔR^2) and corresponding F_{change} values denote the specific improvement of Model 2 over Model 1 in predicting the dependent variable. Toler. = Tolerance; VIF = Variance Inflation Factor. Significant p-values (alpha = .05) depicted in bold typeface.

We hypothesized that performance feedback may be a salient factor that can potentially restore goal-directed control when managing these well-established associations. However, cumulative performance feedback did not break the habits elicited by these familiar stimuli. As seen in the hierarchical multiple regression model, although Condition yielded differential mapping-related NoGo accuracy changes across familiar

and novel conditions, the Feedback regressor was not a significant predictor of outcome-sensitivity.

Secondary measure of outcome-insensitivity: ΔGo_Accuracy

We performed an identical omnibus hierarchical regression using ΔGo_Accuracy, which serves as our secondary measure of outcome-sensitivity. The regression model met the assumptions of normality and homoscedasticity. Multicollinearity tests produced negligible VIFs, confirming the assumption of linearity in the regression model (VIF for all variables < 1.04; see Supplementary Table 5). Two participants were identified as outliers due to a standardized residual values falling outside the predetermined range (Tabachnick & Fidell, 2007). Identical analyses without outlier data produced no substantial changes in the statistical findings below.

Collectively, the linear combination of Age, Gender, and Impulsivity did not significantly predict mapping-related change in Go accuracy: $F(3,96) = 0.31$, $p = .81$. As depicted in Supplementary Table 5, closer examination of the individual regressors revealed no significant role played by any of the controlled variables in Model 1 (all β coefficient *p*'s. > .05; see Supplementary Table 5). In the second step of the regression, the inclusion of the Condition and Feedback regressors explained an additional 17.7% of the variance in Go accuracy impairment: $\beta_{Condition} = -.32$, $p = .001$, $\beta_{Feedback} = .28$, $p = .003$, $\Delta R^2 = .18$, $F_{change}(2,94) = 10.23$, $p < .001$—rendering Model 2 a significant predictor of ΔGo_Accuracy: $F(5,94) = 4.32$, $p = .001$ (see Supplementary Table 5).

Supplemental Table 5. Summary of the Hierarchical Multiple Regression Model for Outcome-Insensitivity as Assayed by ΔGo_Accuracy.

Variable	***Toler.***	***VIF***	***B***	***SE***	***β***	***t***	***sig.***
Model 1							

Age	.99	1.01	-0.11	0.21	-.54	-0.54	.591
Gender	.98	1.02	-1.00	1.38	-.07	-0.73	.469
Impulsivity	.98	1.02	-0.05	0.10	-.05	-0.53	.600
Model 2							
Age	.98	1.02	-0.22	0.20	-.11	-1.13	.261
Gender	.96	1.04	-0.51	1.27	-.04	-0.40	.692
Impulsivity	.98	1.02	-0.04	0.09	-.04	-0.43	.670
Condition	.98	1.02	-4.03	1.19	-.32**	-3.39	**.001**
Feedback	.99	1.01	3.59	1.18	.28**	3.03	**.003**

Model Summary Statistics						
Model	R^2	*F*	*F sig.*	ΔR^2	F_{change}	F_{change} *sig.*
Model 1	.01	0.31	.815			
Model 2	.18	4.32	**.001**	.18	10.23	**<.001**

Note: Top layer of table depicts all regressors included in the hierarchical model and their respective statistics. Bottom layer of table, Model Summary Statistics, depicts the predictive strength of each model. Delta R^2 (ΔR^2) and corresponding F_{change} values denote the specific improvement of Model 2 over Model 1 in predicting the dependent variable. Toler. = Tolerance; VIF = Variance Inflation Factor. Significant p-values (alpha = .05) depicted in bold typeface.

Although these hierarchical regression results regarding ΔGo_Accuracy suggest that cumulative performance feedback has significant predictive strength, the mixed-design ANOVAs in the main text confirm that cumulative performance feedback has a significant effect on Go actions only in the Familiar condition.

Experiment 3 – Omnibus regression to illustrate mapping-related impairment while testing the effect of dual feedback—cumulative performance and monetary feedback

To detect the potential habit-disrupting effects of dual feedback (i.e., paired monetary and cumulative performance feedback), we employed a hierarchical multiple regression model to extract the predictive strengths of the between-group regressors, Condition and Feedback, while controlling for age, gender, and impulsivity. We entered the controlled Age, Gender, and Impulsivity regressors into the first, and Condition and Feedback regressors of interest into the second step of the model to yield an R^2 change (ΔR^2) value that quantifies the contributions of Condition and Feedback. This allowed us to confirm whether mapping-related impairments are predicted specifically by Condition

(i.e., Familiar versus Novel stimuli), and whether dual feedback affected motivational control. A corresponding F_{change} value was derived to confirm whether ΔR^2 reflects a significant change in the model's predictive strength. Participants with DV standardized residual values below -3.3 and above 3.3 were identified as outliers (Tabachnick & Fidell, 2007). In such cases, we performed identical analyses without outlier participants to verify robustness of findings, but only report these excluded analyses if outliers produced substantial changes in statistical significance.

The regression model met the assumptions of normality and homoscedasticity. Multicollinearity tests produced negligible Variance Inflation Factors (VIF), confirming linearity assumptions in the regression (VIF for all variables < 1.15; see Supplementary Table 6).

As hypothesized, Model 1, a linear combination of the controlled variables Age, Gender, and Impulsivity, did not significantly predict outcome-sensitivity: $F(3,96) = 0.12$, $p = .95$, and only explained 0.4% of the variance in ΔNoGo_Accuracy ($R^2 = .004$). Additionally, none of these regressors independently predicted ΔNoGo_Accuracy (all β coefficient p's > .05; see Supplementary Table 6). In the second step of the regression, the inclusion of the Condition and Feedback regressors explained an additional 26.6% of the variance in outcome-sensitivity: $\beta_{Condition} = -.43$, $p < .001$, $\beta_{Feedback} = .28$, $p = .003$, $\Delta R^2 = .27$, $F_{change}(2,94) = 17.16$, $p < .001$—rendering Model 2 a significant predictor of outcome-sensitivity: $F(5,94) = 6.96$, $p < .001$ (see Supplementary Table 6).

Supplemental Table 6. Summary of the Hierarchical Multiple Regression Model for Outcome-Insensitivity as Assayed by ΔNoGo_Accuracy.

Variable	***Toler.***	***VIF***	***B***	***SE***	***β***	***t***	***sig.***
Model 1							
Age	.97	1.03	0.16	0.57	.03	0.29	.775

Gender	.99	1.00	1.31	3.69	.04	0.35	.724
Impulsivity	.97	1.03	0.09	0.22	.04	0.40	.692
Model 2							
Age	.94	1.06	0.24	0.50	.04	0.48	.635
Gender	.87	1.14	0.14	3.41	.004	0.04	.968
Impulsivity	.91	1.09	0.01	0.20	.01	0.08	.939
Condition	.92	1.08	-13.42	2.83	$-.43^{***}$	-4.74	**<.001**
Feedback	.89	1.13	8.76	2.89	$.28^{**}$	3.03	**.003**

Model Summary Statistics

Model	R^2	*F*	*F sig.*	ΔR^2	F_{change}	F_{change} *sig.*
Model 1	.004	0.12	.950			
Model 2	.52	6.96	**<.001**	.27	17.16	**<.001**

Note: Top layer of table depicts all regressors included in the hierarchical model and their respective statistics. Bottom layer of table, Model Summary Statistics, depicts the predictive strength of each model. Delta R^2 (ΔR^2) and corresponding F_{change} values denote the specific improvement of Model 2 over Model 1 in predicting the dependent variable. Toler. = Tolerance; VIF = Variance Inflation Factor. Significant p-values (alpha = .05) depicted in bold typeface.

As hypothesized, the delivery of cumulative performance and monetary feedback disrupted habits (i.e., prevent a significant incongruency-related impairment in NoGo accuracy to familiar lights) and improved goal-directed control (i.e., significantly increase NoGo accuracy to novel stimuli). In other words, these regression data suggest that the differential mapping-related NoGo impairment is replicated in Experiment 3, in that the stimulus condition predicts changes in accuracy, and importantly, dual feedback is able to significantly predict improvements in performance.

Secondary index of outcome-sensitivity: Go Accuracy

We performed identical regression analyses using ΔGo_Accuracy as DV. The regression model met the assumptions of normality and homoscedasticity. Multicollinearity tests produced negligible VIFs, confirming the assumption of linearity in the regression model (VIF for all variables < 1.15). One participant was identified as an outlier due to a standardized residual value less than -3.3 (Tabachnick & Fidell, 2007).

Identical analyses without the outlier data produced no substantial changes in the statistical findings reported below; distinctions are specified where relevant.

Collectively, the linear combination of Age, Gender, and Impulsivity did not significantly predict mapping-related impairments in Go accuracy: $F(3,96) = 1.67$, $p = .18$. As depicted in Supplementary Table 7, Model 1 only explained 5% of the variance ($R^2 = .05$), and of all the controlled variables in Model 1, only the Impulsivity variable significantly predicted ΔGo_Accuracy: $\beta_{Impulsivity} = .21$, $p = .04$ (all other β coefficient p's. > .05). In the second step of our regression model, the inclusion of the Condition and Feedback regressors explained an additional 21.7% of the variance in Go accuracy impairment: $\beta_{Condition} = -.36$, $p < .001$, $\beta_{Feedback} = .28$, $p = .004$, $\Delta R^2 = .21$, $F_{change}(2,94) = 13.11$, $p < .001$—rendering Model 2 a significant predictor of ΔGo_Accuracy: $F(5,94) = 6.50$, $p < .001$ (see Supplementary Table 7). The predictive strength of the Impulsivity variable diminished below significance in Model 2 ($\beta_{Impulsivity} = .17$, $p = .07$). When reanalyzed without outlier data, Impulsivity did not predict ΔGo_Accuracy in either model (both p's > .05). Given its lack of significant contributions in Experiments 1 and 2, and sensitivity to outlier correction in Experiment 3, we refrain from speculating further regarding the robustness of Impulsivity as a predictor.

Supplemental Table 7. Summary of the Hierarchical Multiple Regression Model for Outcome-Insensitivity as Assayed by ΔGo_Accuracy.

Variable	***Toler.***	***VIF***	***B***	***SE***	***β***	***t***	***sig.***
Model 1							
Age	.97	1.03	0.18	0.21	.09	0.87	.384
Gender	.99	1.00	0.78	1.36	.06	0.57	.569
Impulsivity	.97	1.03	0.17	0.08	.21*	2.06	**.042**
Model 2							
Age	.94	1.06	0.20	0.19	.09	1.03	.306
Gender	.87	1.14	0.55	1.30	.04	0.42	.673
Impulsivity	.91	1.09	0.14	0.08	.17	1.84	.069
Condition	.92	1.08	-4.22	1.08	-.36***	-3.92	**<.001**

Feedback	.89	1.13	3.25	1.10	.28**	2.95	**.004**

Model Summary Statistics						
Model	R^2	*F*	*F sig.*	ΔR^2	F_{change}	F_{change} *sig.*
Model 1	.05	1.67	.178			
Model 2	.26	6.50	**<.001**	.21	13.11	**<.001**

Note: Top layer of table depicts all regressors included in the hierarchical model and their respective statistics. Bottom layer of table, Model Summary Statistics, depicts the predictive strength of each model. Delta (Δ) values denote the specific improvement of Model 2 over Model 1 in predicting the dependent variable. Toler. = Tolerance; VIF = Variance Inflation Factor. Significant p-values (alpha = .05) depicted in bold typeface.

The significant Condition and Feedback regressors indicate that like Experiments 1 and 2, Condition (Familiar vs. Novel) differentially yields changes in Go accuracy, and Feedback has a significant improvement effect on Go accuracy. It should be noted that our mixed-design ANOVAs (reported in the main text) reveal a significant feedback effect on Go actions only in the Novel condition. We can thus conclude that dual feedback does not significantly disrupt Go habits, though it does significantly promote goal-directed Go actions.

Chapter 4 Supplement

We uploaded a pre-registration document to Open Science Framework (OSF; https://osf.io) prior to data collection. Barring formatting adjustments, the report below is an unaltered transcript of the planned analyses that can be found on the OSF repository (document URL: https://osf.io/fjcbw).

A. Hypotheses

Description of essential elements

H1a: Participants will exhibit significant NoGo accuracy impairments when managing Go and NoGo signals that are incongruent with the well-established green-go and red-stop associations (Day 1).

H1b: Participants will exhibit flexible performance when managing novel associations, in that NoGo accuracy will be comparable across phases in which purple and blue Go/NoGo signals are presented (Day 1).

H1c: Participants high in ADHD symptom severity, as assayed by ADHD Self Report Scale (ASRS) will exhibit stronger habitual control, as identified by a greater congruency-related impairment when managing familiar stimuli on Day 1.

H1c_alt: Alternatively, ADHD symptom severity of either "inattentive" or "hyperactive" subtypes may specifically predict habitual control. To examine the role of each subtype, participants will be assigned "inattentive" and "hyperactive" subtype scores, derived from Parts A and B of the ASRS survey.

- We have bi-directional hypotheses regarding these subtype variables, in that both inattentive and hyperactive presentations of ADHD may render actions habitual (e.g., hyperactive individuals may indiscriminately respond to Go and NoGo stimuli, and inattentive individuals may have difficulty discriminating Go and NoGo responses due to difficulties attending to the changes in new task demands).

H1d: The driving experience scale will not yield significant correlations with behavioral flexibility as assayed by NoGo accuracy difference across phases (Day 1).

H1e: Order of stimulus block presentation (Familiar block first vs. Novel block first on Day 1) or order of phase presentation (Congruent phase first vs. incongruent phase first) will not differentially affect NoGo accuracy differences (Day 1).

H2a: Participants will exhibit habit disruption following dual feedback (i.e., cumulative performance paired with monetary feedback) receipt on Day 2, such that feedback receipt will serve as a buffer against the significant NoGo accuracy impairment (Day 2).

H2b: Participants high in ADHD symptom severity will benefit less from dual feedback receipt on Day 2, showing heightened NoGo accuracy impairments as a function of ASRS score.

- Similar to H1c_alt, inattentive or hyperactive subtypes may also predict the degree to which dual feedback restores goal-directed control. If H1c_alt reveals subtype-specific results, we expect the same subtype that drives habitual control in the No-Feedback condition to predict accuracy impairment in the Feedback condition.

H2b_alt: Alternatively, participants high in ADHD symptom severity may benefit similarly from dual feedback receipt, rendering dual feedback a salient enough manipulation to restore motivational control even in most-affected participants.

B. Methods

Description of essential elements

Design

IVs:

- Stimulus (Familiar, Novel; within-subject design)
- Phase (Congruent, Incongruent; within-subject design)

DVs:

- NoGo accuracy (errors of commission; main DV of interest)

- Go accuracy (errors of omission; exploratory DV)

Covariates and controlled variables

- ASRS (continuous survey measure; main covariate of interest)
    - Inattentive subtype score (alternative hypothesis testing)
    - Hyperactive subtype score (alternative hypothesis testing)
- Order_Block (Familiar_First vs. Novel_First; between-subject design)
- Order_Phase (Congruent_First vs. Incongruent_First; between-subject design)
- Driving experience scale (continuous scale – months of driving experience; controlled variable)
- Patient (Patient, Non-patient; categorical variable of interest for possible exploratory analyses)

Planned sample

- Exclusion criteria: color-blindness, experience with previous versions of experiment
- Recruitment: undergraduate research subject pool; participation for course credit; study advertisement via online subject pool portal; may recruit paid ADHD patients previously in lab subject database for additional patient vs. non-patient analyses
- We performed a power analysis on data from (Wodushek & Neumann, 2003) – a study in which healthy adults are categorized into high vs. low ADHD symptom groups for inhibitory control comparisons. We extracted effect sizes from correlations between inhibitory control and non-verbal inattention in both symptom severity groups, and averaged the two resulting projected sample sizes.

The averaged sample size needed to reach 80% statistical power was determined to be 105.

- Data collection will terminate upon reaching n=105 (regardless of day 2 attrition rates; day 2 analyses will be performed using available data).

Exclusion criteria

- Trials with reaction times <100 milliseconds (Luce, 1991)
- Standardized residual values outside the -3.3 – 3.3 range (Tabachnick & Fidell, 2007) will be classified as outliers, and analyses excluding these participants will be reported if data removal causes significant changes in results.
- Self-reported color-blindness will disqualify participants from partaking in study.

Procedure

- Day 1 – sequence of experimental events:
    - Informed consent
    - Go/NoGo task instructions
    - Block 1: Familiar or Novel block (order counterbalanced across participants)
        - Phase 1: Congruent or Incongruent phase (order counterbalanced across participants)
        - Contingency reversal instructions
        - Phase 2: same stimulus set as Phase 1, except Go/NoGo contingencies are reversed
    - Block 2: Familiar or Novel block (depending on previous block presented)

        - Phase 1: Congruent or Incongruent phase (order counterbalanced across participants)
        - Contingency reversal instructions
        - Phase 2: same stimulus set as Phase 1, except Go/NoGo contingencies are reversed
    - Surveys: ASRS and Driving Experience Scale administration
- Day 2
    - Go/NoGo task instruction
    - Familiar block
        - Phase 1: Congruent contingencies (Red-NoGo, Green: Go)
        - Dual feedback receipt
        - Contingency reversal instructions
        - Phase 2: Incongruent contingencies (Red:Go, Green:NoGo)
    - Demographic survey, The Creature of Habit Scale (COHS; used in exploratory analysis below), debriefing

Additional manipulation description:

- Stimulus block and phase order counterbalanced across participants.
- Dual feedback manipulation: occurs on Day 2, following Phase 1:
    - Experimenter reveals cumulative performance feedback on screen, leaves room and returns with performance dependent cash bonus with possibility to earn more following next phase (bonus not performance dependent, unbeknownst to participant).

C. Analysis plan

Confirmatory analyses: H1a through H1e: omnibus test encompassing all H1 predictions

Relevant IVs:

- Stimulus (Familiar, Novel)

Relevant DVs:

- NoGo Accuracy Difference (Incongruent minus Congruent Phase NoGo Accuracy; primary DV of interest) and
- Go Accuracy Difference (similar transformation as NoGo Accuracy Difference; exploratory DV)

Covariates and controlled variables of interest:

- ASRS (main covariate of interest, with inattentive and hyperactive sub-scores for alternative hypothesis testing);
- Age,
- Gender,
- Block_Order (Familiar_First, Novel_First),
- Phase_Order (Congruent_First, Incongruent_First),
- Driving Experience

Statistical technique:

- Hierarchical multiple regression
    - (Controlled variables entered into step 1, ASRS entered into step 2, Stimulus entered into step 3).

- Analyses will be repeated with inattentive and hyperactive ASRS sub-scores used as main covariate of interest for alternative hypothesis testing.

Rationale:

- Stimulus variable determines whether accuracy impairment is specific to the Familiar stimulus set.
- NoGo accuracy difference DV ensures the analyses captures behavioral flexibility based on congruency-related impairments.
- Order variables rule out order effects based on block or phase presentation (e.g., training effects)
- ASRS allows for symptom severity and behavioral flexibility hypothesis testing.
    - Inattentive and hyperactive sub-scores will permit testing whether symptom severity specific to these subtypes drives potential behavioral control differences.
- Driving experience as controllable variable ensures effects do not rely on exposure to stimuli solely in a traffic-context (as in, familiar stimuli have well-learned associations that extend beyond driving).
- Demographic variables used to rule out behavioral flexibility variability based on age and gender of participants.

Confirmatory analyses: H2a-H2b_alt:

Relevant IVs:

- Feedback (Feedback, No Feedback)

Relevant DVs:

- NoGo Accuracy Difference (Incongruent minus Congruent Phase NoGo Accuracy; primary DV of interest)
- Go Accuracy (exploratory DV)

Covariates and controlled variables of interest:

- ASRS (main covariate of interest), ASRS subtype score (used in identical analyses if H1c_alt yields subtype-specific information)
- Age,
- Gender,
- Block_Order (Familiar_First, Novel_First),
- Phase_Order (Congruent_First, Incongruent_First),
- Driving Experience

Statistical technique:

- Hierarchical multiple regression
    - (Controlled variables entered into step 1, ASRS and Patient entered into step 2, Feedback entered into step 3).
        - If H1c_alt is confirmed, analyses will be repeated with the ASRS sub-score (inattentive or hyperactive) predictive of accuracy impairment in the No-Feedback condition as main covariate of interest for alternative hypothesis testing.

Rationale:

- Feedback variable determines whether accuracy impairment is specific to the No Feedback condition, potentially illustrating that Feedback prevents habitual control.

- NoGo accuracy difference DV ensures the analyses captures behavioral flexibility based on congruency-related impairments.
- Order variables rule out order effects based on block or phase presentation (e.g., training effects) of day 1 events on day 2 performance.
- ASRS allows for symptom severity and behavioral flexibility hypothesis testing.
- Subtype-specific ASRS sub-scores will permit testing whether symptom severity specific to these subtypes drives potential behavioral control differences in the Feedback condition.
- Driving experience as controllable variable ensures effects do not rely on exposure to stimuli solely in a traffic-context (as in, familiar stimuli have well-learned associations that extend beyond driving).
- Demographic variables used to rule out behavioral flexibility variability based on age and gender of participants.

Planned exploratory analyses:

- Identical analyses will be performed using Go Accuracy Difference DV to further explore behavioral flexibility.
- Go accuracy is used as a supplemental DV due to the high Go:NoGo ratio, expected to promote pre-potent Go responses resulting in high Go accuracy at the expense of relatively lower NoGo accuracy.
- Patient vs. Non-patient analyses
    - ADHD patients may be recruited to: (a) supplement existing dataset with higher ASRS scores for a more complete range of symptom severity, and/or (b) allow clinical vs. sub-clinical ADHD presentation comparisons

of behavioral flexibility. ADHD patient vs. non-patient status will be entered as a categorical variable of interest into the hierarchical regression models.

  - Patients are predicted to exhibit heightened habitual control (pronounced NoGo accuracy difference scores) and to benefit less from dual feedback receipt (significant impairment in NoGo accuracy following contingency reversal on day 2).
  - Alternatively, patients may also exhibit habit disruption following feedback receipt on day 2, serving as an indicator of the efficacy of dual feedback in restoring goal-directed control even in patients with clinical presentations.

- The Creature of Habit Scale (COHS) may be administered depending on time left in session, attrition, and previously collected survey data availability (i.e., if not able to acquire responses in the lab, responses in the pre-screen survey completed by participants at the time of subject pool registration prior to study participation may be used). COHS will be entered into a correlational matrix with Familiar NoGo accuracy difference (incongruent minus congruent) and ASRS score. Thus, we will examine whether habitual tendencies in daily life correlate with habitual control as assayed by congruency-related accuracy impairment in the Familiar condition, as well as ADHD symptom severity.
  - We expect COHS score to be positively correlated with Familiar NoGo Accuracy Difference, such that higher COHS scores will predict

heightened habitual control (i.e., positive correlation as indicated by a significant Pearson's r value.).

- We expect COHS score to significantly correlate with ADHD symptom severity (i.e., positive correlation as indicated by a significant Pearson's r value.)

Recommended elements

- The method of missing data handling: pairwise deletion (day 2 attrition will not eliminate day 1 data)
- Go and NoGo accuracy scores will be transformed into difference scores for regression analyses
- Outliers will be removed from dataset and data will be re-analyzed. If no substantial difference in outlier-included vs. excluded analyses exists, full dataset will be used for reporting results. Otherwise, outlier-removed version will be reported with a description of the deviations occurring due to the outliers.
- Post-hoc ANOVAs and t-tests will be performed where relevant following hierarchical regression models.
    - A repeated measures ANOVA using Stimulus as IV, NoGo Accuracy as DV, Age, Gender, Block_Order, Phase_Order, Driving Experience, and ASRS as covariates.
    - Paired samples t-test across congruent and incongruent phases within each stimulus block (Familiar and Novel)
- Identical analyses using Go accuracy as DV.

Answer the following final questions:

Has data collection begun for this project?

- No, data collection has not begun

The (estimated) start and end dates for this project are (optional): 9/2018 – 05/2019

Any additional comments before I pre-register this project (optional):

- For **H1b**, we expect Novel stimuli to elicit flexible behavioral control, in that no significant impairment occurs as a result of contingency change. This null result ensures that the significant accuracy impairment expected in H1a is indeed due to the Familiar stimulus set. We will confirm the role of stimulus sets by entering Stimulus as a variable in our hierarchical regression model, and as explained above, determine via post-hoc ANOVA and t-tests whether Novel or Familiar stimulus sets elicit accuracy impairments.
- For **H1d and H1e**, we expect the Driving Experience Scale, Order of run and block, and demographic variables to yield null results, in that we will use them as controlled variables in our regression model to rule out contextual (Driving), training (Order), and age/gender (demographic) related effects.
- For **H2a and H2b_alt**, we expect dual feedback receipt to restore goal-directed control, resulting in participants to perform flexibly following contingency change. Although this hypothesis expects a non-significant difference of accuracy across phases, it permits the interrogation of the role of Feedback in affecting behavioral control. The null result in the Feedback condition ensures that the significant accuracy impairment on Day 1 without feedback is indeed due to the

lack of feedback. We will test the significance of Feedback as a candidate for restoring behavioral control by entering it into our hierarchical regression model, and confirm its role via post-hoc ANOVA and t-tests, as explained above.

Chapter 4 Supplement (continued)

Supplemental Table 8. First order correlations between accuracy and individual difference measures.

Variable	**ΔNoGo_Accuracy**	**ΔGo_Accuracy**
ASRS_Inattentive	r = .05, p = .96	r = .10, p = .63
ASRS_Hyperactive	r = .07, p = .96	r = -.06, p = .63
COHS	r = -.15, p = .41	r = -.19, p = .17

Note: p-values have been corrected for multiple comparisons using the Holm-Bonferroni method for each dependent variable.

Our pre-registered hierarchical mixed models did not meet the assumption of multicollinearity due to high correlations between the following regressors: Age and Driving; Phase_Order and Condition_Order; ASRS_Total and ASRS_Inattentive/Hyperactive. We removed the redundant regressors from our analyses (i.e., Age, Condition_Order, and ASRS_Total) that were not integral for our hypotheses to meet the assumption of multicollinearity, and included these corrected analyses in the text. We report below the complete, multicollinear set of regressors for consistency with the pre-registration document. Because of the potentially inflated regression coefficients due to multicollinearity, we refrain from speculating on significant effects in these analyses, and refer readers to the corrected models included in the main text.

Supplemental Table 9. Hierarchical Mixed Model of ADHD Symptomology and Habit Expression: ΔNoGo_Accuracy (As Pre-registered).

Variable	*VIF*	*β*	*t*	*sig.*
Model 1				
Age	6.77	-.25 (.18)	-1.42	.160
Gender	1.10	-.11 (.07)	-1.49	.140
Phase_Order	3.98	.17 (.14)	1.27	.206
Condition_Order	4.11	-.21 (.14)	-1.53	.130
Driving	6.71	.32 (.18)	1.79	.077
Model 2				
Age	7.21	-.27 (.19)	-1.46	.167
Gender	**1.21**	-.09 (.08)	-1.24	.220
Phase_Order	4.21	.17 (.14)	1.24	.218
Condition_Order	4.36	-.22 (.14)	-1.52	.131
Driving	6.99	.33 (.18)	1.79	.077
ASRS_Inattentive	9.00	-.05 (.21)	-0.23	.822
ASRS_Hyperactive	8.74	.02 (.20)	0.08	.934
ASRS_Total	23.53	.03 (.34)	0.10	.917
Diagnosis	1.33	.03 (.08)	0.40	.692
COHS	1.08	-.06 (.07)	-0.77	.442
Model 3				
Age	7.21	-.27 (.18)	-1.54	.128
Gender	**1.21**	-.09 (.07)	-1.30	.197
Phase_Order	4.21	.17 (.13)	1.30	.196
Condition_Order	4.36	-.22 (.14)	-1.60	.113
Driving	6.99	.33 (.17)	1.88	.063
ASRS_Inattentive	9.00	-.05 (.20)	-0.24	.813
ASRS_Hyperactive	8.74	.02 (.19)	0.09	.931
ASRS_Total	23.53	.03 (.32)	0.11	.913
Diagnosis	1.33	.03 (.08)	0.42	.677
COHS	1.08	-.06 (.07)	-0.81	.419
Condition	**1**	**.31 (.07)**	**4.68**	**<.001**
Model 4				
Age	7.21	-.27 (.18)	-1.55	.126
Gender	**1.21**	-.09 (.07)	-1.31	.194
Phase_Order	4.21	.18 (.13)	1.31	.193
Condition_Order	4.36	-.22 (.14)	-1.61	.111
Driving	6.99	.33 (.17)	1.89	.062
ASRS_Inattentive	17.56	-.05 (.20)	-0.24	.812
ASRS_Hyperactive	17.22	.02 (.19)	0.09	.931
ASRS_Total	46.31	.03 (.32)	0.11	.912
Diagnosis	2.38	.03 (.08)	0.42	.675
COHS	2.13	-.06 (.07)	-0.82	.417
Condition	**64.97**	**.31 (.07)**	**4.70**	**<.001**
ASRS_Inattentive x Condition	91.11	.17 (.19)	0.89	.376
ASRS_Hyperactive x Condition	71.66	.15 (.19)	0.76	.448
ASRS_Total x Condition	265.78	-.35(.31)	-1.12	.265
Diagnosis x Condition	2.16	.10 (.07)	1.54	.126
COHS x Condition	57.37	.12 (.07)	1.74	.085

Model Comparisons

Model	R^2	*Log likel.*	χ^2	χ^2 *sig.*	ΔR^2
Model 1	.05	81.55			
Model 2	.05	81.99	0.89	.971	<.01
Model 3	**.15**	**92.99**	**22.00**	**<.001**	**.10**
Model 4	.18	96.83	7.69	.174	.03

Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Bottom layer of table, Model Comparisons, depicts the predictive strength of each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface.

Supplemental Table 10. Hierarchical Mixed Model of ADHD Symptomology and Habit Disruption: ΔNoGo_Accuracy (As Pre-registered).

Variable	*VIF*	*β*	*t*	*sig.*
Model 1				
Age	6.77	.29 (.18)	1.61	.111
Gender	1.10	.04 (.07)	0.60	.550
Phase_Order	**3.98**	**.30 (.14)**	**2.20**	**.030**
Condition_Order	4.11	-.23 (.14)	-1.67	.098
Driving	6.71	-.28 (.18)	-1.54	.126
Model 2				
Age	7.21	.32 (.18)	1.72	.088
Gender	1.21	.05 (.08)	0.62	.537
Phase_Order	**4.21**	**.30 (.14)**	**2.14**	**.035**
Condition_Order	4.36	-.25 (.14)	-1.71	.091
Driving	6.99	-.26 (.18)	-1.45	.151
ASRS_Inattentive	9.00	.05 (.21)	0.24	.809
ASRS_Hyperactive	8.74	.28 (.20)	1.37	.174
ASRS_Total	23.53	-.27 (.33)	-0.80	.423
Diagnosis	1.33	-.04 (.08)	-0.53	.595
COHS	1.08	-.08 (.07)	-1.18	.239
Model 3				
Age	7.21	.32 (.18)	1.80	.076
Gender	1.21	.05 (.07)	0.65	.520
Phase_Order	**4.21**	**.30 (.14)**	**2.23**	**.028**
Condition_Order	4.36	-.25 (.14)	-1.78	.078
Driving	6.99	-.26 (.17)	-1.51	.135
ASRS_Inattentive	9.00	.05 (.20)	0.25	.801
ASRS_Hyperactive	8.74	.28 (.20)	1.43	.157
ASRS_Total	23.53	-.27 (.32)	-0.84	.404
Diagnosis	1.33	-.04 (.08)	-0.55	.580
COHS	1.08	-.08 (.07)	-1.23	.221
Feedback	**1**	**-.28 (.07)**	**-4.18**	**<.001**
Model 4				
Age	7.21	.32 (.18)	1.79	.076
Gender	1.21	.05 (.07)	0.64	.521
Phase_Order	**4.21**	**.30 (.14)**	**2.22**	**.029**
Condition_Order	4.36	-.25 (.14)	-1.78	.079
Driving	6.99	-.26 (.18)	-1.50	.136
ASRS_Inattentive	17.56	.05 (.20)	0.25	.802
ASRS_Hyperactive	17.22	.28 (.20)	1.42	.158
ASRS_Total	46.31	-.27 (.32)	-0.84	.405
Diagnosis	2.38	-.04 (.08)	-0.55	.581
COHS	2.13	-.08 (.07)	-1.23	.222
Feedback	**64.97**	**-.28 (.07)**	**-4.17**	**<.001**
ASRS_Inattentive x Feedback	91.11	-.23 (.19)	-1.20	.232
ASRS_Hyperactive x Feedback	71.66	-.30 (.19)	-1.53	.128
ASRS_Total x Feedback	265.78	.52 (.32)	1.62	.108
Diagnosis x Feedback	2.16	.01 (.07)	0.21	.837
COHS x Feedback	57.37	-.06 (.07)	-0.87	.388

Model Comparisons

Model	R^2	*Log likel.*	χ^2	χ^2 *sig.*	ΔR^2
Model 1	.04	75.55			
Model 2	.06	77.72	4.35	.501	.02
Model 3	**.14**	**86.58**	**17.73**	**<.001**	**.08**
Model 4	.16	88.82	4.47	.484	.02

Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Bottom layer of table, Model Comparisons, depicts the predictive strength of each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface.

Supplemental Table 11. Hierarchical Mixed Model of ADHD Symptomology and Habit Expression: ΔGo_Accuracy (As Pre-registered).

Variable	*VIF*	*β*	*t*	*sig.*
Model 1				
Age	6.77	-.07 (.18)	-0.41	.686
Gender	1.10	.03 (.07)	0.39	.693
Phase_Order	3.98	.04 (.14)	0.31	.756
Condition_Order	4.11	-.01 (.14)	-0.07	.941
Driving	6.71	.12 (.18)	0.66	.512
Model 2				
Age	7.21	-.19 (.19)	-1.04	.301
Gender	1.21	.08 (.08)	1.06	.291
Phase_Order	4.21	.05 (.14)	0.39	.697
Condition_Order	4.36	-.04 (.14)	-0.29	.773
Driving	6.99	.15 (.18)	0.81	.421
ASRS_Inattentive	**9.00**	**-.47 (.21)**	**-2.25**	**.027**
ASRS_Hyperactive	**8.74**	**-.48 (.20)**	**-2.37**	**.020**
ASRS_Total	**23.53**	**.78 (.33)**	**2.33**	**.022**
Diagnosis	1.33	.15 (.08)	1.85	.067
COHS	1.08	-.06 (.07)	-0.85	.397
Model 3				
Age	7.21	-.19 (.18)	-1.05	.294
Gender	1.21	.08 (.07)	1.08	.284
Phase_Order	4.21	.05 (.14)	0.40	.692
Condition_Order	4.36	-.04 (.14)	-0.29	.770
Driving	6.99	.15 (.18)	0.82	.414
ASRS_Inattentive	**9.00**	**-.47 (.20)**	**-2.28**	**.025**
ASRS_Hyperactive	**8.74**	**-.48 (.20)**	**-2.41**	**.018**
ASRS_Total	**23.53**	**.78 (.33)**	**2.37**	**.020**
Diagnosis	1.33	.15 (.08)	1.88	.063
COHS	1.08	-.06 (.07)	-0.86	.390
Condition	**1**	**.18 (.07)**	**2.64**	**<.009**
Model 4				
Age	7.21	-.19 (.18)	-1.06	.290
Gender	**1.21**	.08 (.07)	1.08	.281
Phase_Order	4.21	.05 (.14)	0.40	.690
Condition_Order	4.36	-.04 (.14)	-0.30	.768
Driving	6.99	.15 (.18)	0.83	.410
ASRS_Inattentive	17.56	-.47 (.20)	-2.29	.024
ASRS_Hyperactive	17.22	-.48 (.20)	-2.42	.017
ASRS_Total	46.31	.78 (.33)	2.38	.019
Diagnosis	2.38	.15 (.08)	1.89	.061
COHS	2.13	-.06 (.07)	-0.87	.386
Condition	**64.97**	**.18 (.07)**	**2.66**	**<.009**
ASRS_Inattentive x Condition	91.11	-.22 (.20)	-1.09	.279
ASRS_Hyperactive x Condition	71.66	-.02 (.20)	-0.12	.906
ASRS_Total x Condition	265.78	.14 (.32)	0.43	.666
Diagnosis x Condition	**2.16**	**.15 (.07)**	**2.19**	**.031**
COHS x Condition	57.37	.05 (.07)	0.77	.442

Model Comparisons

Model	R^2	*Log likel.*	χ^2	χ^2 *sig.*	ΔR^2
Model 1	.01	195.33			
Model 2	**.06**	**201.12**	**11.59**	**.041**	**.05**
Model 3	**.09**	**204.76**	**7.28**	**.007**	**.03**
odel 4	.13	208.98	8.44	.133	.04

Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Bottom layer of table, Model Comparisons, depicts the predictive strength of each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface.

Supplemental Table 12. Hierarchical Mixed Model of ADHD Symptomology and Habit Disruption: ΔGo_Accuracy (As Pre-registered).

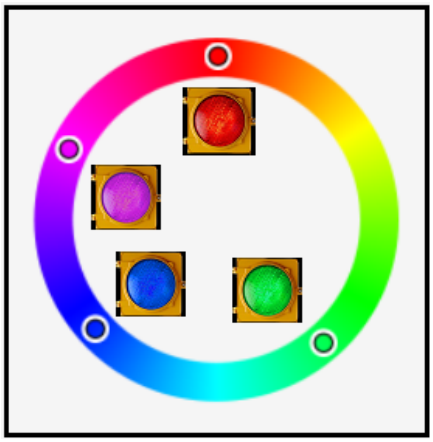
Variable	*VIF*	*β*	*t*	*sig.*
Model 1				
Age	6.77	-.06 (.18)	-0.33	.737
Gender	1.10	.01 (.07)	0.09	.928
Phase_Order	**3.98**	**.36 (.14)**	**2.54**	**.012**
Condition_Order	4.11	-.26 (.14)	-1.83	.071
Driving	6.71	.03 (.18)	0.18	.855
Model 2				
Age	7.21	-.16 (.18)	-0.89	.377
Gender	1.21	.01 (.07)	0.20	.845
Phase_Order	**4.21**	**.34 (.14)**	**2.44**	**.016**
Condition_Order	4.36	-.25 (.14)	-1.73	.086
Driving	6.99	.11 (.18)	0.60	.550
ASRS_Inattentive	9.00	-.23 (.21)	-1.13	.261
ASRS_Hyperactive	**8.74**	**-.42 (.20)**	**-2.07**	**.041**
ASRS_Total	23.53	-.52 (.33)	1.57	.120
Diagnosis	1.33	-.02 (.08)	-0.30	.766
COHS	1.08	-.08 (.07)	-1.17	.245
Model 3				
Age	7.21	-.16 (.18)	-0.89	.377
Gender	1.21	.01 (.08)	0.20	.845
Phase_Order	**4.21**	**.34 (.14)**	**2.44**	**.016**
Condition_Order	4.36	-.25 (.14)	-1.73	.086
Driving	6.99	.11 (.18)	0.60	.550
ASRS_Inattentive	9.00	-.23 (.21)	-1.13	.261
ASRS_Hyperactive	**8.74**	**-.42 (.20)**	**-2.07**	**.041**
ASRS_Total	23.53	-.52 (.33)	1.57	.121
Diagnosis	1.33	-.02 (.08)	-0.30	.766
COHS	1.08	-.08 (.07)	-1.17	.245
Feedback	**1**	**-.24 (.06)**	**-3.70**	**<.001**
Model 4				
Age	7.21	-.16 (.19)	-0.88	.383
Gender	1.21	.01 (.08)	0.19	.847
Phase_Order	**4.21**	**.34 (.14)**	**2.40**	**.018**
Condition_Order	4.36	-.25 (.15)	-1.71	.090
Driving	6.99	.11 (.18)	0.59	.556
ASRS_Inattentive	17.56	-.23 (.21)	-1.12	.267
ASRS_Hyperactive	**17.22**	**-.42 (.21)**	**-2.04**	**.044**
ASRS_Total	46.31	-.52 (.34)	1.55	.125
Diagnosis	2.38	-.02 (.08)	-0.29	.769
COHS	2.13	-.08 (.07)	-1.15	.251
Feedback	**64.97**	**-.24 (.06)**	**-3.84**	**<.001**
ASRS_Inattentive x Feedback	91.11	-.12 (.18)	-0.68	.496
ASRS_Hyperactive x Feedback	71.66	-.34 (.18)	-1.86	.065
ASRS_Total x Feedback	265.78	.48 (.30)	1.62	.107
Diagnosis x Feedback	2.16	<-.01 (.06)	<-0.01	.996
COHS x Feedback	57.37	-.10 (.06)	-1.57	.120

Model Comparisons

Model	R^2	*Log likel.*	χ^2	χ^2 *sig.*	ΔR^2
Model 1	.04	318.23			
Model 2	.07	321.76	7.06	.216	.03
Model 3	**.13**	**328.57**	**13.63**	**<.001**	**.06**
Model 4	.16	333.64	10.13	.072	.04

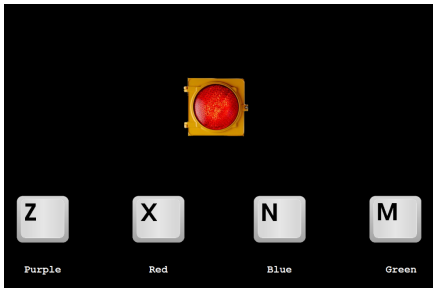
Note: Top layer of table depicts all regressors included in the hierarchical model. Standard errors are given in parentheses. Bottom layer of table, Model Comparisons, depicts the predictive strength of each model, as compared to its previous step. VIF = Variance Inflation Factor. Log likel. = Log likelihood. Significant p-values depicted in bold typeface.

Chapter 5 Supplement



Supplemental Figure 1. Task stimuli superimposed onto the color wheel. Hex color codes for select pixels in each traffic light stimuli are as follows: Red: #ff0000, Purple: #d12fdf, Blue: #1d47f5, Green: #03d547. Color wheel adapted from color-hex (https://www.color-hex.com/color-wheel/).

In Chapters 3 and 4, we found that colors with existing Go and NoGo associations (red-stop, green-go) elicited outcome-insensitive actions, while newly learned Go/NoGo associations (purple-go, blue-stop) were labile. A potential concern could be that because purple is located between red and blue on the color wheel, these perceptually similar color stimuli may be difficult to discern (see Supplemental Figure 1). Ideally, a yellow stimulus that renders all four colors comparably distant on the color wheel could circumvent such an issue. However, yellow may also be associated with ambiguous representations, as it commonly serves as a signal for "caution" and "slow". Alternatively, ascertaining that the blue and purple stimuli are comparably discernible would solidify the notion that the outcome-insensitivity effect is driven not by perceptual differences, but differences in stimulus-response strength.

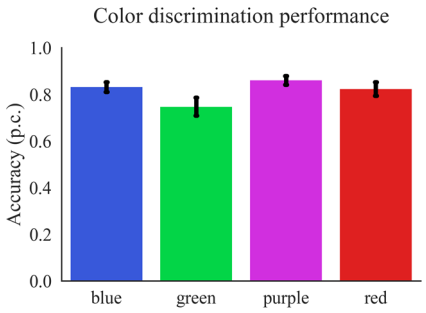


Supplemental Figure 2. Color discrimination task. Participants identify stimulus colors using index and middle fingers. A legend mapping each color to a button remains on the screen throughout the experiment.

To this end, we created a color discrimination paradigm to further validate the well-learned habit task utilized in Chapters 3 and 4. Participants underwent a color identification task, which consisted of the traffic light stimuli from our well-learned habit task (n = 24). Participants were instructed to identify the color

that was displayed in each trial as quickly and accurately as possible. In each trial, a traffic light stimulus appeared, and participants used one of the four response options to select the corresponding color before stimulus offset. Each stimulus remained on the screen for 800 ms. We reasoned that because the Go/NoGo task demanded two color-response mappings at a time with a 400 ms response window, a color discrimination task with four color-response mappings in each trial would warrant an 800 ms response window. Similar to the Go/NoGo task, an inter-trial interval randomly varying between 1200 and 2400 ms separated each trial. The colors were mapped to Z, X, N, and M on the keyboard, and participants used their index and middle fingers to select their responses. A color-button mapping legend remained at the bottom of the screen throughout the experiment, and these mappings were counterbalanced across subjects (see Supplemental Figure 2). The task comprised 80 trials (20 of each color).

The color discrimination performance of our participants is depicted in Supplemental Figure 3. We aimed to reveal whether red, blue, and purple stimuli were comparably distinct, in that participants identified these colors with similar accuracy. We performed a one-way repeated-measures ANOVA across all four colors, and post-hoc t-tests between colors of interest. We found a significant difference in color discrimination between all four stimuli per our ANOVA results, $F(3,69) = 4.10$, $p = .010$. However, post-hoc t-tests showed that there is no significant difference in color discrimination



Supplemental Figure 3. Color discrimination performance. We found no significant differences in color discrimination performance between the stimuli that were thought to be perceptually similar, in that participants did not differ in performance between blue and purple (p = .287), or red and purple (p = .284). p.c.: proportion correct.

between blue and purple, $t(23) = -1.09$, $p = .287$, or red and purple, $t(23) = 1.10$, $p = .284$. Possibly, the significant ANOVA is driven by the identification of the green stimulus. It should be noted that discrimination performance for all colors is well above chance. The green stimulus elicited the lowest discrimination performance at 74.6% correctly identified, compared to chance performance of 25%. Given the comparable performance between the stimuli of interest (purple-red and blue-purple), we can conclude that our results in Chapter 3 and 4 are not due to a difficulty in discerning these colors.