Fall 10-1-2021

# Specifying and Targeting Cognitive-Affective Dysfunctions in Antisocial Individuals

Allison Mary Stuppy-Sullivan
*Yale University Graduate School of Arts and Sciences*, alli.stuppy@gmail.com

Abstract

Specifying and Targeting Cognitive-Affective Dysfunctions in Antisocial Individuals

Allison M. Stuppy-Sullivan

2021

Antisocial behavior includes a wide range of behaviors that violate social norms, from criminal acts to substance misuse. The adverse consequences of antisocial behavior produce a great physical and emotional burden on perpetrators, victims, and family members. This burden is not addressed adequately, with incarceration being the most common intervention for antisocial behavior. When individuals who chronically engage in antisocial behavior are offered therapeutic treatments, the majority neither complete nor benefit from them. One reason existing treatments do not fully address antisocial behavior is because they do not consider or target cognitive-affective dysfunctions driving such behavior, and mechanistic research, to date, does not adequately characterize these cognitive-affective dysfunctions. The present dissertation consists of three studies that refine accounts of cognitive-affective dysfunctions contributing to antisocial behavior and demonstrate how targeting identified dysfunctions can improve cognition and behavior in chronically antisocial individuals. More specifically, Study 1 examines how reward features impact perception, executive functioning, and risk-based decision-making in antisocial individuals. Study 2 examines how reward information is integrated during effort-based decision-making in antisocial individuals, and how negative affect impacts this integration. Finally, Study 3 tests a novel cognitive remediation training package designed to address cognitive-affective dysfunctions in antisocial individuals. Across the three studies in this dissertation, findings highlight that cognitive-affective

dysfunctions related to antisocial behavior reflect difficulty integrating information in specific affectively charged circumstances, and call for a less pessimistic view about treatment for antisocial behavior, and the burden it produces, when these dysfunctions are considered.

# Specifying and Targeting Cognitive-Affective Dysfunctions in Antisocial Individuals

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Allison M. Stuppy-Sullivan

Dissertation Director: Arielle R. Baskin-Sommers, Ph.D.

Dissertation Chair: BJ Casey, Ph.D.

December 2021

## Table of Contents

**Acknowledgements**

I would first like to thank my adviser, Arielle-Baskin Sommers, for the enormous amount of time and effort she has spent helping me become a better scientist and clinician. I truly cannot thank her enough for letting me join the lab, and for creating a mentorship space where I (and many other lucky students) could learn and grow, pursuing meaningful projects at a level I could not have dreamed of before starting graduate school. Most individuals do not express gratitude towards those who sent them to prison, but I can truly say I cherish the opportunities she afforded me to spend hours in confinement and reflect on antisocial behavior in the last several years. She has seamlessly served as a needed source of both support and challenge in my development, and I am forever grateful. Moving forward, I hope to follow Arielle's example of practicing tremendous empathy—even for those in society who receive the least—and making painstaking efforts to ensure that strong science delivers needed changes.

I would also like to thank the remainder of my committee members—BJ Casey, Jutta Joormann, Ifat Levy, and Matthew Steinfeld—for modeling scientific curiosity and offering their time and feedback to improve my work as a clinical scientist. Whether I was in class with BJ or Jutta discussing exciting ways science could extend across levels of analysis, reading thoughtful comments BJ and Ifat made on my prior research projects, or discussing multivariate constructs and strategies relevant to treating my patients in outpatient substance use treatment under Matt's supervision, I have consistently learned and been inspired by each of them over the last several years.

One member of my committee, the late Kathy Carroll, is unfortunately not with us today. She made the treatment study included in this dissertation possible, with her lab's

infrastructure, her thoughtful input, and her incredible vision of the science of behavior change. It has been difficult to picture submitting my dissertation without her. I am grateful for the many moments I could see her in action, beyond those that involved providing insights on my projects. She was constantly brainstorming ways to increase the accessibility of evidence-based therapies, to communicate genuine care and appreciation for study participants, and to improve the wellbeing of members of the New Haven community at large. I hope to follow in her example of using science and warmth to improve the lives of others.

I must also thank the many individuals who helped me become the clinician I am today. In addition to Matt Steinfeld, I would like to thank Mary O'Brien, Emily Cooney, and Seth Axelrod for their helpful insights, teachings, and feedback as I have developed as a therapist. Each of them has modeled, taught, or encouraged techniques and reflections that have made me a better clinician, scientist, and person. Their mentorship has positively impacted me and all of the community members and veterans with whom I have worked in the last five years. I know for a fact that through their direct care and mentorship, they have saved numerous lives and will continue to save more.

Data collection for this dissertation would not have been possible without the many lab members and research assistants who helped run behavioral tasks and conducted clinical interviews, including Scott Tillem, Grace Brennan, Suzy Estrada, Ariel Chen, Yudilyn Jaramillio, Elizabeth Zordani, Emily Beckford, Haley Mitchell-Adams, Josh Hayden, Cole Rianda, and Erika Lopez. I also thank members of the MoD Lab and Psychotherapy Development Center more broadly for their support of these projects and the larger research community they fostered.

I would also like to thank staff at the Connecticut Department of Correction, the Substance Abuse Treatment Unit of the Connecticut Mental Health Center, and the Yale New Haven Hospital Intensive Outpatient Program for allowing this research to take place in their facilities. This work also would not have been possible without the many participants who agreed to take part in these studies despite the discomfort that may have arisen when discussing past experiences, completing neuropsychological measures, and playing computer games that required effort and prompted negative affect. My genuine hope is that these discomforts are offset by the insights they provided, which can ultimately help us better understand and treat antisocial individuals.

Finally, I am grateful to my family for their continuous love, affirmation, support, and entertainment. My mother, Peggy Stuppy, has set an incredible example of what it means to validate others, work hard on others' behalf, and instill an unwavering sense of hope and positivity. My father, John Stuppy, has taught me how to think scientifically and creatively throughout all walks of life, to dream big, and to keep moving despite obstacles. My siblings—Melia, Thomas, and John Stuppy—have helped me develop both a thick skin and an enormous capacity to consider others, providing daily benefit to me personally and professionally. My husband, Mike Sullivan, has been an endless source of joy, validation, and inspiration, and I could not be luckier receiving the love, encouragement, and adventure he brings me. My son, Jude, fills my days with reward and joy. To him, and all who have helped me take care of him; thank you.

# Chapter 1: General Introduction

Antisocial behavior is a heterogeneous construct that includes a wide range of behaviors from criminal acts to substance misuse. The adverse consequences of antisocial behavior produce a great physical and emotional burden on perpetrators, victims, and family members. Moreover, the economic burden of antisocial behavior is immense, with estimated annual costs of over $3 trillion in the United States alone (Baskin-Sommers, Curtin, & Newman, 2015; Federal Bureau of Investigation, 2016; National Institute on Drug Abuse, 2017). The enormous financial and psychological toll associated with antisocial behavior highlights the importance of conducting programmatic research related to its etiology and translating findings into targeted treatments.

Broadly speaking, antisocial individuals that engage in chronic and severe behavior can be classified clinically based on distinct psychopathologies using the Diagnostic and Statistical Manual (DSM-5; American Psychiatric Association, 2013). One such diagnosis is Antisocial Personality Disorder (APD), which reflects repeated violations of social norms, impulsivity, aggression, and irresponsibility. APD is represented in approximately 2-4% of the general population (Compton, Conway, Stinson, Colliver, & Grant, 2005; Fisher & Hany, 2019; Glenn, Johnson, & Raine, 2013), but over 50% of incarcerated individuals (Black, 2015). Another diagnostic category reflective of antisociality is Substance Use Disorders (SUDs), which are characterized by cognitive, behavioral, and physiological symptoms reflecting continued substance use despite significant substance-related problems. Twelve-month prevalence for SUDs is 7.4% in the general population (Substance Abuse and Mental Health Services Administration, 2019) but 65% in the US prison population (National Institute on Drug Abuse, 2020). APD and SUDs frequently co-occur: 40 to 50% of males with SUDs have

APD, and 90% of individuals diagnosed with APD have SUDs (Forrest, 1994; Gerstley, Alterman, McLellan, & Woody, 1990; Messina, Wish, & Nemes, 1999; Tims, De Leon, & Jainchill, 1994). The systematic co-occurrence of SUDs and APD, and their shared common cognitive-affective profiles, suggest these traits reflect an underlying externalizing spectrum (Krueger et al., 2005).

Most previous research on individuals with antisocial psychopathology highlights diminished executive functioning and aberrant decision-making as core cognitive dysfunctions contributing to their problematic behavior (Blair, 2001; the dorsolateral prefrontal cortex executive functioning in APD model; Dolan & Park, 2002; the cognitive–neurobehaivoral model of alcoholism; Giancola & Moss, 1998; the impaired response inhibition and salience attribution model; iRISA; Goldstein & Volkow, 2002; Goldstein & Volkow, 2011; the P3 amplitude reduction in externalizing model; Iacono, Malone, & McGue, 2003; the prefrontal–amygdala–striatal model of APD; Raine, 2018; the 3-pathway model; Verheul, 2001). Additionally, motivational and affective sensitivities are apparent, with evidence that individuals with APD or who are high on externalizing are prone to excessive reward seeking across a variety of reward-related cues (e.g., monetary, substance; Baskin-Sommers & Newman, 2013; Volkow & Li, 2004). These motivational and affective sensitivities undermine cognitive processes even further (Bechara, 2005; Bickel, Jarmolowicz, Mueller, Gatchalian, & McClure, 2012; Bickel et al., 2007; Jentsch & Taylor, 1999; McClure & Bickel, 2014). According to dual-systems models, antisocial individuals fail to self-regulate because they have not only a diminished ability to exercise control (i.e., they have difficulty applying metaphorical “brakes”), but also have abnormally strong appetitive urges (i.e., they have

a heavy foot on a metaphorical “gas pedal;" Bell & McBride, 2010; Buckholtz, 2015; Casey & Caudle, 2013) that further undermine control. Following these models, there is strong evidence that antisocial individuals have decreased capacities to employ executive functions and make cost-benefit decisions, particularly in rewarding contexts (Buckholtz, 2015; Buckholtz, Karmarkar, Ye, Brennan, & Baskin-Sommers, 2017; Dolan, 2012; Garcia-Villamisar, Dattilo, & Garcia-Martinez, 2017; Mazas, Finn, & Steinmetz, 2000; Morgan & Lilienfeld, 2000; Ogilvie, Stewart, Chan, & Shum, 2011; Patrick, Durbin, & Moser, 2012; Petry, 2002; Rowe, 1997). While there is substantial evidence that antisocial individuals display executive dysfunctions, abnormal decision-making, and aberrant reward sensitivies, extant research is limited in numerous ways.

First, while several studies indicate executive dysfunction, abnormal decision-making, and reward sensitivity among antisocial individuals, several other studies provide counterevidence of these abnormalities. For example, many studies report no overall declines in executive functions (e.g., Baskin-Sommers et al., 2014; Maes & Brazil, 2013; Smith, Mattick, Jamadar, & Iredale, 2014; Stevens, Kaplan, & Hesselbrock, 2003), and report instances of intact cost-benefit decision-making by antisocial individuals (e.g., De Brito, Viding, Kumari, Blackwood, & Hodgins, 2013; Gregory et al., 2015; Swogger, Walsh, Lejuez, & Kosson, 2010), challenging the idea that such deficits are at the core of antisocial behavior. Further, several studies indicate antisocial individuals can perform complex cognitive tasks in the presence of rewards (e.g., Charles-Walsh, Upton, & Hester, 2016; Swogger et al., 2010), undermining the notion that reward sensitivity affects their executive functioning and decision-making. In light of the inconsistent findings across studies, it is possible that the links between antisociality,

executive functioning deficits, abnormal cost-benefit decision-making, and reward sensitivity, while intuitive, are overstated.

Second, previous research also is limited in that most studies lack specificity regarding discrete executive functioning and decision-making processes (Ogilvie et al., 2011; Raine & Scerbo, 1991). Executive functioning refers to a diverse range of interdependent cognitive processes that control and coordinate subprocesses of cognition, guiding complex tasks and goal-directed behavior (Miyake et al., 2000), including working memory, future planning, set shifting, and response inhibition (Burgess, 1997; Jurado & Rosselli, 2007; Maes & Brazil, 2013; Miyake et al., 2000; Ogilvie et al., 2011; Royall et al., 2002; Salthouse, 2005; Smith & Jonides, 1999; Stuss & Knight, 2002). Similarly, cost-benefit decision-making can be parsed based on the distinct type of cost associated with a given choice option (e.g. risk, delay, ambiguity; effort; Rudebeck, Walton, Smyth, Bannerman, & Rushworth, 2006). Many previous studies do not differentiate among these component processes, and may thus oversimplify cognitive dysfunctions, and the role of reward in exacerbating such dysfunctions, among antisocial individuals. Understanding how cognitive and reward processes interact is important and, ultimately, mapping interactions among these processes may best determine the specific processes underlying antisocial behavior. The current literature, though, falls short in its specification and integration of the vast cognitive and reward-related processes that affect antisocial individuals.

Finally, reward sensitivity is not the only affective abnormality among antisocial individuals (Baskin-Sommers & Newman, 2013). Evidence suggests that abnormal cognitive performance in distressing contexts also is common for individuals with

antisocial psychopathology (e.g., Daughters et al., 2005; Daughters, Sargeant, Bornovalova, Gratz, & Lejuez, 2008), and some models of antisociality propose negative emotionality as a core feature (in the prefrontal–amygdala–striatal model of APD; Raine, 2018; the 3-pathway model; Verheul, 2001). Further research is necessary to characterize how a wider range of affective states impact cognition in antisocial individuals.

Overall, a more thorough examination of the specific cognitive-affective dysfunctions in antisocial individuals may lead to a better understanding of the mechanisms underlying and maintaining their behavior. Moreover, greater specificity in accounting for the cognitive-affective dysfunctions present among antisocial individuals is likely to inform better treatment approaches for individuals who are largely considered to be treatment resistant and prone to relapse and recidivism (National Institute on Drug Abuse, 2018; Raine, 2018). Across three studies, this dissertation incorporates methodologies from experimental science to refine cognitive-affective conceptualizations of antisocial behavior for a prevalent, chronic, and severe subset of antisocial individuals and to illustrate how leveraging such conceptualizations can improve treatment efficacy in this population.

Study 1 utilizes a well-established reward priming paradigm across three cognitive tasks to identify specific circumstances in which cognition and reward interact to disrupt behavior in a sample of incarcerated individuals diagnosed with APD. Study 2 utilizes a well-established decision-making task to characterize how individuals with APD symptoms in the community incorporate cost-benefit information when choosing to expend effort, and how this information integration is impacted by negative affect. Finally, Study 3 is a pilot randomized controlled trial testing a novel cognitive

remediation training package designed to address antisociality-linked cognitive-affective dysfunctions in a sample of antisocial individuals enrolled in outpatient substance use treatment. Taken together, this set of studies aims to elucidate the cognitive-affective dysfunctions that contribute to antisocial behavior and demonstrate how targeting such dysfunctions can meaningfully improve cognition and behavior.

## General Introduction: References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2015). Altering the cognitive-affective dysfunctions of psychopathic and externalizing offender subtypes with cognitive remediation. *Clin Psychol Sci, 3*(1), 45-57. doi:10.1177/2167702614560744

Baskin-Sommers, A. R., Krusemark, E. A., Curtin, J. J., Lee, C., Vujnovich, A., & Newman, J. P. (2014). The impact of cognitive control, incentives, and working memory load on the P3 responses of externalizing prisoners. *Biological Psychology, 96*, 86-93. doi:10.1016/j.biopsycho.2013.12.005

Baskin-Sommers, A. R., & Newman, J. P. (2013). Differentiating the Cognition-Emotion Interactions That Characterize Psychopathy versus Externalizing. In *Cognition and emotion* (pp. 501-520). New York: Guilford Press.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. *Nature neuroscience, 8*(11), 1458-1463.

Bell, C. C., & McBride, D. F. (2010). Affect regulation and prevention of risky behaviors. *JAMA, 304*(5), 565-566.

Bickel, W. K., Jarmolowicz, D. P., Mueller, E. T., Gatchalian, K. M., & McClure, S. M. (2012). Are executive function and impulsivity antipodes? A conceptual reconstruction with special reference to addiction. *Psychopharmacology, 221*(3), 361-387.

Bickel, W. K., Miller, M. L., Yi, R., Kowal, B. P., Lindquist, D. M., & Pitcock, J. A. (2007). Behavioral and neuroeconomics of drug addiction: competing neural systems and temporal discounting processes. *Drug and alcohol dependence, 90*, S85-S91.

Black, D. W. (2015). The natural history of antisocial personality disorder. *The Canadian Journal of Psychiatry, 60*(7), 309-314.

Blair, R. J. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery & Psychiatry, 71*(6), 727-731.

Buckholtz, J. W. (2015). Social norms, self-control, and the value of antisocial behavior. *Current Opinion in Behavioral Sciences, 3*, 122-129. doi:10.1016/j.cobeha.2015.03.004

Buckholtz, J. W., Karmarkar, U., Ye, S., Brennan, G. M., & Baskin-Sommers, A. (2017). Blunted ambiguity aversion during cost-benefit decisions in antisocial individuals. *Scientific reports, 7*(1), 1-9.

Burgess, P. W. (1997). Theory and methodology in executive function research. *Methodology of frontal and executive function*, 81-116.

Casey, B., & Caudle, K. (2013). The teenage brain: Self control. *Current Directions in Psychological Science, 22*(2), 82-87.

Charles-Walsh, K., Upton, D. J., & Hester, R. (2016). Examining the interaction between cognitive control and reward sensitivity in substance use dependence. *Drug and alcohol dependence, 166*, 235-242.

Compton, W. M., Conway, K. P., Stinson, F. S., Colliver, J. D., & Grant, B. F. (2005). Prevalence, correlates, and comorbidity of DSM-IV antisocial personality syndromes and alcohol and specific drug use disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Journal of Clinical Psychiatry, 66*(6), 677-685.

Daughters, S. B., Lejuez, C., Bornovalova, M. A., Kahler, C. W., Strong, D. R., & Brown, R. A. (2005). Distress tolerance as a predictor of early treatment dropout in a residential substance abuse treatment facility. *Journal of Abnormal Psychology, 114*(4), 729-734.

Daughters, S. B., Sargeant, M. N., Bornovalova, M. A., Gratz, K. L., & Lejuez, C. (2008). The relationship between distress tolerance and antisocial personality disorder among male inner-city treatment seeking substance users. *Journal of Personality Disorders, 22*(5), 509-524.

De Brito, S. A., Viding, E., Kumari, V., Blackwood, N., & Hodgins, S. (2013). Cool and hot executive function impairments in violent offenders with antisocial personality disorder with and without psychopathy. *PLoS One, 8*(6), e65566.

Dolan, M. (2012). The neuropsychology of prefrontal function in antisocial personality disordered offenders with varying degrees of psychopathy. *Psychological Medicine, 42*(8), 1715-1725.

Dolan, M., & Park, I. (2002). The neuropsychology of antisocial personality disorder. *Psychological Medicine, 32*(3), 417-427. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11989987

Federal Bureau of Investigation. (2016). Crime in the U.S. 2015. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015

Fisher, K. A., & Hany, M. (2019). Antisocial Personality Disorder.

Forrest, G. G. (1994). *Chemical dependency and antisocial personality disorder: Psychotherapy and assessment strategies*: Psychology Press.

Garcia-Villamisar, D., Dattilo, J., & Garcia-Martinez, M. (2017). Executive functioning in people with personality disorders. *Current opinion in psychiatry, 30*(1), 36-44.

Gerstley, L. J., Alterman, A. I., McLellan, A. T., & Woody, G. E. (1990). Antisocial personality disorder in patients with substance abuse disorders: A problematic diagnosis? *The American journal of psychiatry, 147*(2), 173.

Giancola, P. R., & Moss, H. B. (1998). Executive cognitive functioning in alcohol use disorders. In *Recent developments in alcoholism* (pp. 227-251): Springer.

Glenn, A. L., Johnson, A. K., & Raine, A. (2013). Antisocial personality disorder: a current review. *Current Psychiatry Reports, 15*(12), 1-8. doi:10.1007/s11920-013-0427-7

Goldstein, R. Z., & Volkow, N. D. (2002). Drug Addiction and its Underlying Neurological Basis: Neuroimaging Evidence for the Involvement of the Frontal Cortex. *American Journal of Psychiatry, 159*(10), 1642-1652.

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews Neuroscience, 12*(11), 652-669.

Gregory, S., Blair, R. J., Simmons, A., Kumari, V., Hodgins, S., & Blackwood, N. (2015). Punishment and psychopathy: a case-control functional MRI investigation

of reinforcement learning in violent antisocial personality disordered men. *The Lancet Psychiatry, 2*(2), 153-160.

Iacono, W. G., Malone, S. M., & McGue, M. (2003). Substance use disorders, externalizing psychopathology, and P300 event-related potential amplitude. *International Journal of Psychophysiology, 48*(2), 147-178. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12763572

Jentsch, J. D., & Taylor, J. R. (1999). Impulsivity resulting from frontostriatal dysfunction in drug abuse: implications for the control of behavior by reward-related stimuli. *Psychopharmacology, 146*(4), 373-390.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review, 17*(3), 213-233.

Maes, J. H., & Brazil, I. A. (2013). No clear evidence for a positive association between the interpersonal-affective aspects of psychopathy and executive functioning. *Psychiatry Research, 210*(3), 1265-1274. doi:10.1016/j.psychres.2013.09.028

Mazas, C. A., Finn, P. R., & Steinmetz, J. E. (2000). Decision-making biases, antisocial personality, and early-onset alcoholism. *Alcoholism: Clinical and Experimental Research, 24*(7), 1036-1040.

McClure, S. M., & Bickel, W. K. (2014). A dual-systems perspective on addiction: contributions from neuroimaging and cognitive training. *Annals of the New York Academy of Sciences, 1327*(1), 62-78.

Messina, N. P., Wish, E. D., & Nemes, S. (1999). Therapeutic community treatment for substance abusers with antisocial personality disorder. *Journal of Substance Abuse Treatment, 17*(1), 121-128.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.

Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review, 20*(1), 113-136. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10660831

National Institute on Drug Abuse. (2017, April 21, 2017). Trends & Statistics. Retrieved from https://www.drugabuse.gov/related-topics/trends-statistics

National Institute on Drug Abuse. (2018). Principles of Drug Addiction Treatment: A Research-Based Guide. Third Edition. Retrieved from https://www.drugabuse.gov/publications/principles-drug-addiction-treatment-research-based-guide-third-edition on 2018, January 20

National Institute on Drug Abuse. (2020). Criminal Justice DrugFacts. Retrieved from https://www.drugabuse.gov/publications/drugfacts/criminal-justice

Ogilvie, J. M., Stewart, A. L., Chan, R. C. K., & Shum, D. H. K. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology, 49*(4), 44.

Patrick, C. J., Durbin, C. E., & Moser, J. S. (2012). Reconceptualizing antisocial deviance in neurobehavioral terms. *Development and Psychopathology, 24*(3), 1047-1071. doi:10.1017/s0954579412000533

Petry, N. M. (2002). Discounting of delayed rewards in substance abusers: relationship to antisocial personality disorder. *Psychopharmacology, 162*(4), 425-432.

Raine, A. (2018). Antisocial Personality as a Neurodevelopmental Disorder. *Annual Review of Clinical Psychology*(0).

Raine, A., & Scerbo, A. (1991). Biological theories of violence. In *Neuropsychology of aggression* (pp. 1-25): Springer.

Rowe, D. C. (1997). Are parents to blame? A look at the antisocial personalities. *Psychological Inquiry, 8*(3), 251-260.

Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., . . . Coffey, C. E. (2002). Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. *The Journal of neuropsychiatry and clinical neurosciences, 14*(4), 377-405.

Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. (2006). Separate neural pathways process different decision costs. *Nature neuroscience, 9*(9), 1161-1168.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*(4), 532.

Smith, J. L., Mattick, R. P., Jamadar, S. D., & Iredale, J. M. (2014). Deficits in behavioural inhibition in substance abuse and addiction: A meta-analysis. *Drug and alcohol dependence, 145*, 1-33. doi:10.1016/j.drugalcdep.2014.08.009

Stevens, M. C., Kaplan, R. F., & Hesselbrock, V. M. (2003). Executive–cognitive functioning in the development of antisocial personality disorder. *Addictive behaviors, 28*(2), 285-300.

Stuss, D. T., & Knight, R. T. (2002). *Principles of frontal lobe function*: Oxford University Press.

Substance Abuse and Mental Health Services Administration. (2019). Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health. Retrieved from https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf

Swogger, M. T., Walsh, Z., Lejuez, C. W., & Kosson, D. S. (2010). Psychopathy and Risk Taking among Jailed Inmates. *Criminal Justice and Behavior, 37*(4), 439-452. doi:10.1177/0093854810361617

Tims, F. M., De Leon, G., & Jainchill, N. (1994). *Therapeutic community: Advances in research and application* (F. M. Tims, G. De Leon, & N. Jainchill Eds.): US Department of Health and Human Services, Public Health Service, National Institutes of Health, Natonal Instutitute on Drug Abuse.

Verheul, R. (2001). Co-morbidity of personality disorders in individuals with substance use disorders. *European Psychiatry, 16*(5), 274-282.

Volkow, N. D., & Li, T.-K. (2004). Drug addiction: the neurobiology of behaviour gone awry. *Nature Reviews Neuroscience, 5*(12), 963-970.

## Chapter 2: Study 1

## Evaluating dysfunction in cognition and reward among offenders with Antisocial Personality Disorder

Published in:

Stuppy-Sullivan, A., & Baskin-Sommers, A. (2019). Evaluating dysfunction in cognition and reward among offenders with antisocial personality disorder. *Personality Disorders: Theory, Research, and Treatment, 10*(5), 416.

**Abstract**

Antisocial Personality Disorder (APD) is a costly clinical condition. Previous studies identify executive dysfunction and reward sensitivity as factors contributing to APD. However, empirical evidence supporting the role of these factors in APD is mixed. The present study aimed to identify and specify APD-related dysfunction in cognitive and reward factors. In a sample of incarcerated males (N=116), we administered three tasks targeting distinct cognitive (perception, executive functioning, and probabilistic decision-making) and reward (magnitude and consciousness) factors. APD was associated with impaired perception when high magnitude rewards were at stake, regardless of reward consciousness. APD was associated with worse executive functioning during conscious high rewards, as well as worse inhibition during high rewards when working memory demands were high. There was no APD-related performance difference during probabilistic decision-making. These findings expose the multifaceted nature of cognitive-affective dysfunction in APD, highlighting the importance of systematic research and providing insight into treatment targets.

*Keywords:* antisocial personality disorder; cognition; reward; perception; executive functioning

Antisocial personality disorder (APD) is a costly clinical condition associated with a persistent pattern of social, legal, and moral norm violations (American Psychiatric Association, 2013). The prevalence of APD is markedly elevated in incarcerated offenders, with evidence that rates of APD are approximately 13 times higher in prisoners compared with the general population (Compton, Conway, Stinson, Colliver, & Grant, 2005; Fazel & Danesh, 2002). Individuals with APD represent a particularly high-risk subtype of offenders, committing higher rates of violent and nonviolent crimes, obtaining diagnoses of severe forms of substance use disorders (Brennan, Stuppy-Sullivan, Brazil, & Baskin-Sommers, 2017), and having increased mortality rates (National Institute for Health Clinical Excellence, 2009) compared with individuals without APD. Despite the significance of APD as a driver of costly behavior, we still know relatively little about the cognitive and affective factors underlying the disorder. This is due, in part, to the failure of previous research to systematically specify factors of cognition and affect that are disrupted in APD.

Based on existing research, executive dysfunction and reward hypersensitivity emerge as possible candidate factors implicated in the pathogenesis of APD. Across studies and meta-analyses, individuals with APD show deficits in many components of executive functioning (Dolan, 2012; Garcia-Villamisar, Dattilo, & GarciaMartinez, 2017; Morgan & Lilienfeld, 2000; Ogilvie, Stewart, Chan, & Shum, 2011; Patrick, Durbin, & Moser, 2012; Rowe, 1997) including inhibition (Barkataki et al., 2008; Chamberlain, Derbyshire, Leppink, & Grant, 2016; De Brito, Viding, Kumari, Blackwood, & Hodgins, 2013; Dolan & Park, 2002; Rubio et al., 2007; Swann, Lijffijt, Lane, Steinberg, & Moeller, 2009; Zeier, Baskin-Sommers, Hiatt Racer, & Newman, 2012), planning (Dolan

& Park, 2002), working memory (Dolan & Park, 2002), and set shifting (Dolan & Park, 2002). Moreover, the extant literature describes individuals with APD as exemplars of a dominant reward-based system (Quay, 1993). Empirical evidence indicates that individuals with APD are hypersensitive to rewards (Raine, 2018; Völlm et al., 2010), resulting in their strong desire for immediate rewards (Petry, 2002), even when their reward-driven behavior is accompanied by negative consequences (Mazas, Finn, & Steinmetz, 2000). Together, research provides strong support for the purported relationships among APD, executive dysfunction, and reward hypersensitivity. Moreover, the nature of these relationships seems intuitive, given that individuals with APD repeatedly display behaviors reflecting a failure to inhibit urges (e.g., fighting and crime), and they often do so in pursuit of rewards (e.g., to obtain other's property in the case of theft or to achieve a "high" from substance use).

Although the work noted earlier suggests diminished executive functioning and heightened reward sensitivity among individuals with APD, the exact cognitive-affective factors at issue remain somewhat underspecified. First, take cognition. It is clear from decades of research that cognition contains multiple separable factors, including perception (supporting encoding and early attention), executive functioning (discrete functions supporting complex tasks and goal-directed behavior [e.g., monitoring, updating, suppressing competing memory representations in working memory, planning, set shifting, and inhibition]) and decision-making (supporting the evaluation of and choices between alternative actions; Burgess, 1997; Jurado & Rosselli, 2007; Maes & Brazil, 2013; Miyake et al., 2000; Ogilvie et al., 2011; Purves et al., 2008; Royall et al., 2002; Salthouse, 2005; Smith & Jonides, 1999; Stuss & Knight, 2002). In general,

cognition can be impacted in a variety of ways based on these factors, and dysfunction associated with any one of these factors may disrupt processing associated with other factors. With these cognitive factors in mind, close examination of the existing research on APD and executive functioning actually highlights that some tasks used to tap executive functioning also manipulate perception (e.g., Cambridge gambling task [CGT]) or decision-making (e.g., Iowa gambling task [IGT]; Snyder, Miyake, & Hankin, 2015).

For example, some research of executive dysfunction in APD reports poor performance among individuals with APD during tasks like the IGT (Bechara, Damasio, Damasio, & Anderson, 1994; Gansler, Jerram, Vannorsdall, & Schretlen, 2011; Mazas et al., 2000) and the CGT (De Brito et al., 2013; Rogers et al., 1999). The IGT, though, examines several cognitive factors within executive functions (e.g., set shifting, planning, and working memory) and decision-making (e.g., value-based learning, reversal learning, and risk-aversion; De Brito & Hodgins, 2009). Likewise, on the CGT, performance "quality" depends not only on executive functions and decision-making but also the perceptual capability of an individual to discern among various visual stimuli. With multiple cognitive factors assessed during tasks like the IGT and CGT, it is unclear whether poor performance for those with APD reflects executive dysfunction or whether abnormal perception, decision-making, or an interaction among these cognitive factors promotes dysfunction in these individuals. Moreover, even studies using purportedly "purer" measures of executive function, such as set-shifting or planning tasks, do not support the claim that individuals with APD show fundamental deficits in executive functions (Chamberlain et al., 2016; Crowell, Kieffer, Kugeares, & Vanderploeg, 2003; De Brito et al., 2013; Maes & Brazil, 2013; Stevens, Kaplan, & Hesselbrock, 2003).

Across multiple types of executive functioning tasks, individuals with APD tend to show dysfunction under high cognitive load (e.g., when planning several steps and maintaining complex stimuli over long periods of time; De Brito et al., 2013; Dolan & Park, 2002) and during inhibition of prepotent responses (De Brito et al., 2013; Dolan & Park, 2002). At this point, extant literature in APD has not provided a clear picture of dysfunction, either in terms of specific executive functions or with regard to cognitive dysfunction more broadly.

Second, reward also can be subdivided into multiple separable factors. Common factors include reward magnitude (the amount of reward available; Beilock, 2007; Berridge, 2004; Knutson, Adams, Fong, & Hommer, 2001; Knutson, Taylor, Kaufman, Peterson, & Glover, 2005; Mobbs et al., 2009; Robbins & Everitt, 1996; Schultz, 2006) and reward consciousness (the degree to which awareness of reward information can bias behavior; Berridge, Robinson, & Aldridge, 2009; Berridge & Winkielman, 2003; Bijleveld, Custers, & Aarts, 2009; Pessiglione et al., 2008; van Gaal & Lamme, 2012; Zedelius et al., 2014). Each of these alone or combined can contribute to an individual's reward sensitivity. Different laboratory paradigms use controlled manipulations of these factors to quantify their common and unique impact on an individual's behavior. This approach allows researchers to clarify and contextualize cognitive and reward abnormalities. Unfortunately, many tasks selected for research on reward sensitivity in APD conflate multiple reward factors or subtly assess components of reward without fully manipulating those components, making it difficult to know which components of reward processing, if any, are affected in APD.

As an example, risky decision-making tasks conflate reward magnitude and reward probability by exclusively pairing low magnitude rewards with high probabilities and high magnitude rewards with low probabilities, such that the influence of magnitude or probability cannot be disentangled (e.g., IGT and balloon analogue risk task; Lejuez et al., 2002). Unfortunately, because many decision-making tasks do not use systematic reward magnitude manipulations, it is unclear whether the observed reward sensitivity in individuals with APD reflects sensitivity to reward magnitude, reward probability, or a combination of these reward features (Dolan & Park, 2002; Mazas et al., 2000; Swogger, Walsh, Lejuez, & Kosson, 2010). Another example relates to how reward consciousness has been a factor of reward noted in research on APD, but not examined systematically. Individuals with APD appear reward hypersensitive when they are not consciously aware of reward information (e.g., they display a "decision bias" during early trials of the IGT when they are unaware of reward contingencies, Mazas et al., 2000; they show abnormal neural responding during a rewarded color discrimination task in which they are not aware of when or how much rewards are available, Völlm et al., 2010). By contrast, individuals with APD do not show reward hypersensitivity when contingencies are more explicit (e.g., during later trials of the IGT when they are more aware of reward outcomes and probabilities associated with each option, Mazas et al., 2000; during the balloon analogue risk task when they are aware of the gains and losses at stake for taking risks, Swogger et al., 2010). These findings suggest that for individuals with APD, an unconscious bias toward reward information may disrupt behavior, but also that conscious awareness (i.e., explicit presentation) of reward may regulate their behavior. However, the tasks used in these studies do not implement validated reward

consciousness manipulations and only examine unconscious reward processing indirectly (i.e., after rewards are obtained). Thus, across studies, the common tasks used to assess reward sensitivity in APD do not systematically manipulate reward magnitude or reward consciousness. The observed reward sensitivity in individuals with APD may reflect sensitivity to rewards of specific magnitudes, an unconscious bias to rewards, or sensitivity to rewards more broadly.

Although a substantial body of research highlights abnormalities in cognition and reward in APD, a closer examination of a largely equivocal literature highlights a need for more systematic research isolating specific factors. The goal of the present study is to systematically assess factors of cognition and reward to identify specific dysfunction(s) in individuals with APD. In a sample of incarcerated offenders, we administer three cognitive tasks and simultaneously manipulate reward using well-established manipulations. Given the strong association between APD and executive functions documented in previous research, one task selected is a modified *n*-back task. This is an executive function task that combines elements from the most widely used tasks for assessing the cognitive factors that are most robustly associated with APD: inhibition (e.g., go/no-go and stop-signal tasks; Congdon et al., 2012) and working memory (Owen, McMillan, Laird, & Bullmore, 2005). Another task is a visual search task to assess individual ability to identify target stimuli among distractors (Wolfe, 1998) because successful performance on many go/no-go and working memory tasks, including the *n*-back task, involves discerning among visual stimuli. Finally, a probabilistic gambling task is used because a multitude of studies purported to assess executive functioning in APD often target decision-making processes, with the most equivocal decision-making

findings in APD related to decision-making under risk (Buckholtz, Karmarkar, Ye, Brennan, & Baskin-Sommers, 2017; De Brito et al., 2013; Mazas et al., 2000). The selected decision-making task includes two-choice decisions with explicit outcome values and probabilities, removing any need for reward learning or contingency updating, which are often conflated in tasks intended to measure decision-making under risk (De Brito et al., 2013; Dunn, Dalgleish, & Lawrence, 2006). All participants complete the perceptual visual search task first, followed by the executive function *n*-back task and the decision-making probabilistic gambling task.[1]

During each of these tasks, reward magnitude (low vs. high) and awareness of reward information (conscious vs. unconscious) is manipulated.[2] First, reward magnitude is selected because decades of research across disciplines document its importance as a modulator of behavior among healthy individuals (Beilock, 2007; Berridge, 2004; Mobbs et al., 2009; Pessiglione et al., 2007; Robbins & Everitt, 1996; Schultz, 2006; Zedelius, Veling, & Aarts, 2011; Zedelius et al., 2014), and some studies suggest individuals with

[1] Cognition is a multidimensional construct that can be divided into separable but interrelated factors. The selected tasks follow examples in existing literature that manipulate only one aspect of cognition at a time. For example, the perceptual visual search task only taxes encoding; in the executive function *n*-back task, inhibition and working memory are manipulated, and perceptual load is held constant across trials; and, probabilistic decision-making varies across the decision-making probability gambling task, whereas perceptual load and working memory are constant. Thus, although it is expected that several cognitive factors are represented in some of the tasks, each task manipulates only one cognitive factor at a time. This represents a departure from the tasks previously used to examine cognitive functioning in APD, which often manipulate multiple cognitive factors simultaneously.

[2] As noted earlier, reward sensitivity can be multifaceted (Berridge et al., 2009), with reward magnitude and consciousness being just two of several established reward factors (see also reward probability and reward delay; Schultz, 2006). For the present study, reward magnitude and consciousness are selected because across studies of reward sensitivity in APD, different levels of reward magnitude and reward consciousness appear to be associated with divergent findings, and well-established methods manipulating these factors are available to examine the impact of these factors directly and simultaneously.

APD respond strongly to reward magnitude manipulations (Mazas et al., 2000). Second, reward consciousness is selected based on recent cognitive neuroscience evidence suggesting individual variability in sensitivity to conscious and unconscious rewards (Bustin, Quoidbach, Hansenne, & Capa, 2012; Zedelius et al., 2014) that also may impact the quality of executive functioning (Capa & Bouquet, 2018; Capa, Bustin, Cleeremans, & Hansenne, 2011), a factor of cognition purportedly important in the pathogenesis of APD. Although there are hints that reward magnitude and consciousness influence reward sensitivity *across* APD studies, neither reward magnitude nor reward consciousness is varied systematically *within* any current study of reward sensitivity in APD. Thus, in the present study, reward magnitude and consciousness are manipulated systematically and simultaneously (i.e., fully crossed across all trials of the three cognitive tasks) to isolate the impact of these factors on individuals with APD.

Together, this design allows us to examine components of cognition and reward processing, and how they interact, to identify vulnerabilities related to APD. Current conceptualizations of APD cite a vastly mixed literature concerning cognitive and reward processes, and it is essential that we refine our understanding of these processes to identify the most likely circumstances in which cognition and reward result in antisocial behavior.

## Method

### Participants

Participants were 116 men from a maximum-security correctional facility, between the ages of 18 and 75; with an IQ greater than 70, a reading level of at least

fourth grade, no clinical diagnoses of schizophrenia, bipolar disorder, or psychosis; who were not currently using psychotropic medications; and who did not have medical problems that could impact comprehension.[3] Participants completed a diagnostic interview to assess criteria for APD on one visit and the three laboratory tasks on a second visit (see Table 1 for sample characteristics and Methods in the Supplemental Material for full details). All participants were provided written informed consent according to the procedures set forth by the Yale University Institutional Review Board.

**Tasks**

**Masked reward cues[4].** Before each trial in the three tasks, the point value at stake for the trial was displayed using a modified reward-masking paradigm (Figure 1; Bijleveld et al.,
2009). Point values were low (1 point) or high (10 points), noted by blocked digits (01 and 10, respectively). These reward cues were displayed either consciously (i.e., for a duration that is consciously perceivable, 300 ms) or unconsciously (i.e., 30ms; see Methods in the Supplementary Materials for full details).

[3] *A priori* power analyses based on previous studies on related topics (e.g., individual differences in perception, *n*-back, cost-benefit decision-making) were conducted using G*Power statistical software (Faul, Erdfelder, Lang, & Buchner, 2007). Power analyses indicated that a sample size of 98-128 participants would result in sufficient (80%) power to detect a moderate effect for the omnibus interactions between repeated measures within-subjects task conditions and a between-subjects variable.

[4]To ensure that participants were unable to consciously perceive the 30-ms unconscious reward cues, subliminality was tested in a random subset of the participants after completion of the three main tasks. A total of 25 participants were presented with 20 masked reward cues, in the same manner as in the unconscious (30 ms) reward cue used throughout the study. Participants indicated the value of each presented reward cue (01 or 10). Performance for discriminating between the unconscious reward cues was no better than chance, $M_{accuracy}$ = .52, SD = .09, t(24) = 1.28, p = .212, 95% $CI_{diff}$ [-0.01, 0.06].

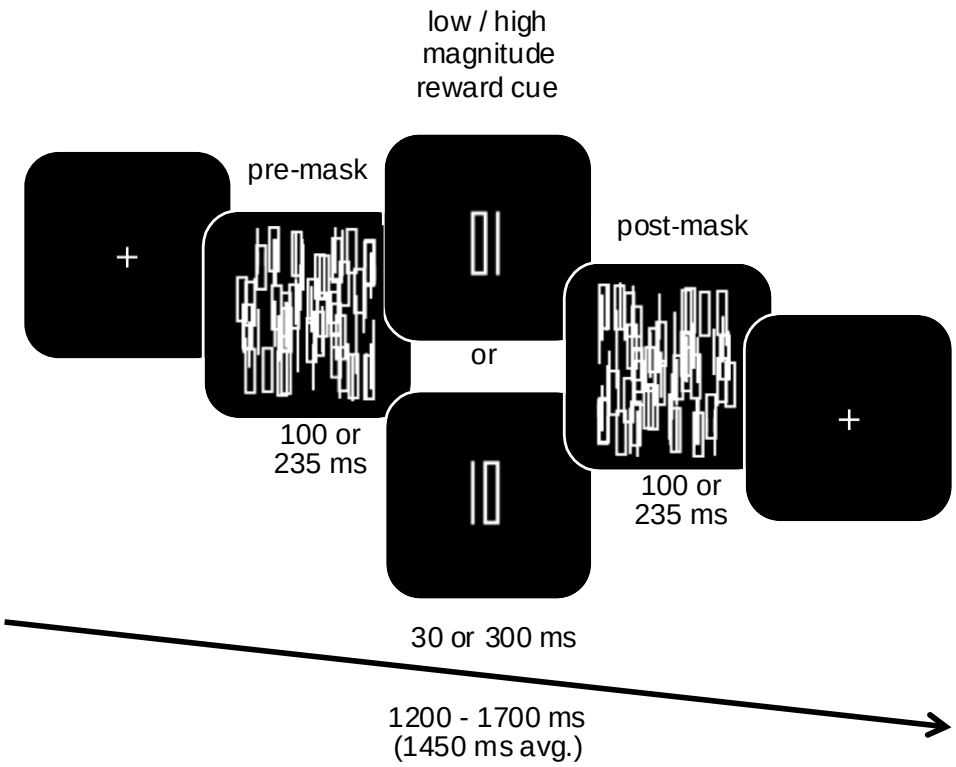


**Figure 1. Reward mask procedure.** Each masked reward cue lasted 500ms and was preceded and followed by fixation (total procedure lasts 1200-1700ms, 1450ms on average). Reward cues were either "01" for low rewards or "10" for high rewards, with blocked edges. Before and after each reward cue, a mask consisting of overlapping 0's and 1's with blocked edges was presented. For unconscious cues, masks were presented for 235ms before and after cues, with reward cues presented for 30ms. For conscious cues, masks were presented for 100ms before and after cues, with reward cues presented for 300ms. Participants were told that reward information will be presented to inform them of the reward value at stake for each trial, and that this information might be difficult to see at times.

**Visual search task.** For the perception task, a modified version of a visual search task was used (Kristjánsson, Sigurjónsdóttir, & Driver, 2010; Figure 2A). During the task, participants viewed a series of displays with three colored diamonds. Participants were instructed to search for the oddly colored diamond, either a red target among two green distractors or vice versa. Participants indicated (by button press) whether the oddly colored diamond had a notch missing at the top or the bottom of the shape. Because performance for this task may include changes in speed or accuracy, an inverse efficiency score (IES; mean response time for correct responses divided by percentage of correct

responses) was calculated for each participant (see Methods in the Supplemental Material for full details).

***n*-back task.** For the executive functioning task, we used a modified version of the *n*-back task (Figure 2B; Baskin-Sommers et al., 2014; Pochon et al., 2002). During the task, participants viewed a series of letters. Participants were instructed to monitor the letters and respond with a button press if the preceding letter in the *n*-back position was different from the current letter (e.g., a mismatch trial). Participants were instructed to withhold their response when the preceding letter matched the current stimulus (e.g., a match trial). The majority of trials were mismatch trials (80%), whereas match trials were infrequent (occurring 20% of the time). The task also included a manipulation of working memory load. In the low-load (1-back) condition, participants were instructed to determine whether the currently presented letter matched the immediately preceding letter in the sequence. In the high-load (2-back) condition, participants were required to monitor and maintain the stimulus information in working memory to determine whether the letter stimulus two positions earlier matched the current letter. For each participant, accuracy on the task was calculated (see Methods in the Supplemental Material for full details).

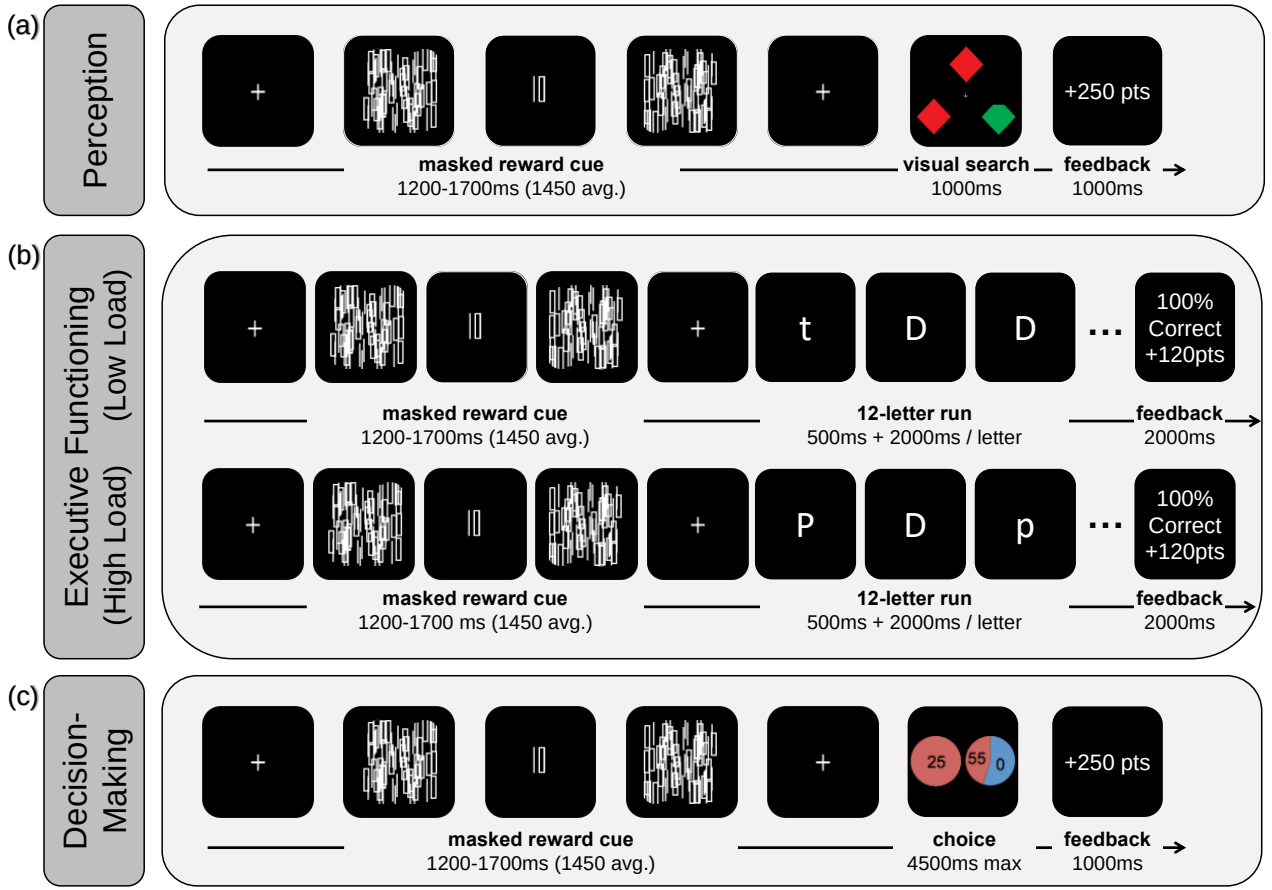


**Figure 2. Example of a trial in each of the three tasks.** (a). For the perception task, each trial began with a masked reward cue presented between fixation crosses (1450ms on average). Participants were presented with a visual search display and asked to respond by indicating via button press whether a colored diamond had a notch missing from the top or bottom of it (1000ms). Participants were then provided with feedback (1000ms) about whether they responded correctly within the time limit and how many points they earned for doing so. (b). For the executive functioning task, each trial began with a masked reward cue presented between fixation crosses. Participants were presented with a series of letters (500ms/each, with a 2000ms delay between letters). Participants were asked to press a button for each letter, unless the letter matched the letter immediately before it in a 1-back trial (first row in middle) or the letter two before it in a 2-back trial (second row in middle). Following a run of 12 letters (i.e., trial), participants were provided with feedback (2000ms) about the percentage of correct responses and how many points they earned for the run. **(c).** For the probabilistic decision-making task, each trial began with a masked reward cue presented between fixation crosses. Participants were presented with two circles showing a choice between a small certain reward and a larger probabilistic reward (4500ms). Participants chose one of the two options via button press and were informed how many points they earned (1000ms).

**Gambling task.** To assess probabilistic decision-making, a gambling task was used to examine risk-taking behavior (modified gain conditions from Voon et al., 2006; Figure 2C). During the task, participants viewed a series of two circles (i.e., gamble options). Participants were instructed to make a choice between one of two gamble options: a "sure" and a "risky" option. Participants were to press the right button for the option on the right of the screen and left button for the option on the left of the screen. For each participant, the percentage of "risky" choices was calculated (see Methods in the Supplemental Material for full details).

## Results

### Visual Search Task

First, we analyzed IES in a General Linear Model (GLM) with reward magnitude (low vs. high) and reward consciousness (conscious vs. unconscious) as within-subjects categorical factors and IQ (z-scored) as a continuous covariate[5]. Consistent with previous research, there was a significant main effect for reward consciousness, $F(1, 114) = 30.68$, $p < .001$, $\eta^2 = .21$, 95% CI: [.11, .31], indicating higher IES (worse speed-accuracy) for unconscious compared to conscious reward cues (Bijleveld et al., 2009; Bijleveld, Custers, & Aarts, 2010; Pessiglione et al., 2007; Zedelius et al., 2011). There was no

[5] IQ was included as a covariate in analyses for all task variables (visual search, *n*-back, and gambling), as IQ was related to both task performance and APD. Moreover, in additional analyses we examined the specificity of the effects reported in the text by including related disinhibitory psychopathologies (i.e., substance use disorders or psychopathy). The visual search and *n*-back by APD effects remain the same. The only exception is when controlling for substance use disorders the *n*-back reward magnitude by reward consciousness by APD effect becomes non-significant. Finally, we examined whether the number of symptoms of APD (i.e., continuous count of CD and adult antisocial symptoms) predicted the same effects reported in the text. When using a continuous count of APD symptoms, the visual search and *n*-back by APD effects remain the same. Therefore, the APD-related effects reported for the visual search and *n*-back tasks hold up for a continuous measure of antisocial behavior and are specific to APD.

main effect for reward magnitude ($p$ = .425) or an interaction between reward magnitude and consciousness ($p$ = .129)

Second, the association between encoding and APD was examined by including APD (present vs. absent) in the GLM as a between-subjects categorical factor. There was a significant interaction between reward magnitude and APD, $F(1, 113) = 7.11$, $p$ = .009, $\eta^2$ = .06, 95% CI: [.01, .14] (see Figure 3). For individuals with APD, there was a significant effect of reward magnitude, such that individuals with APD showed higher IES (worse speed-accuracy) for high compared to low reward cues during visual search ($p$ = .015, $\eta^2$ = .05, 95% CI: [.01, .13). For individuals without APD, there was no effect of reward magnitude ($p$ = .195). Neither the main effect for APD nor any other APD by task interaction was significant (all $p$'s > .25).

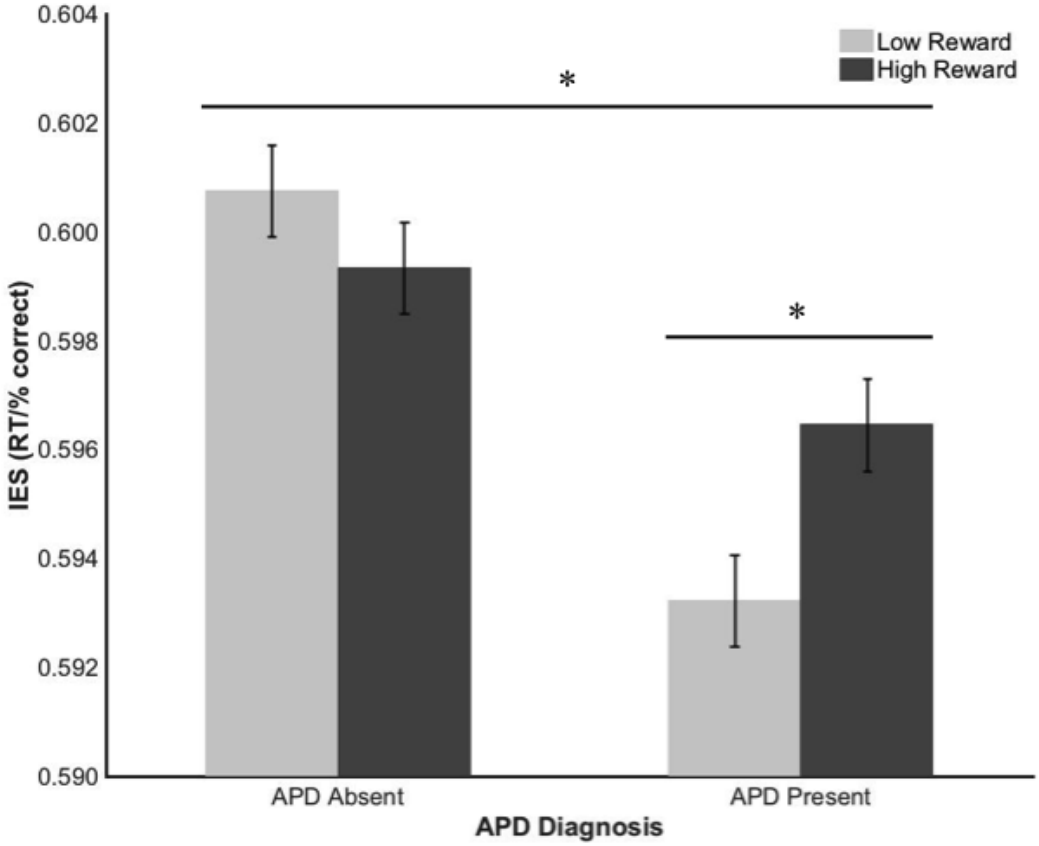


**Figure 3. Perception and APD Effects.** There was a significant interaction between reward magnitude and APD. Individuals with APD showed higher IES (worse speed-accuracy) for high compared to low reward cues during visual search, whereas individuals without APD were unaffected by reward magnitude. Error bars represent 1 within-subjects standard error.

### *n*-back Task

First, accuracy on the *n*-back task was examined using a GLM with reward magnitude (low vs. high), consciousness (conscious vs. unconscious), trial type (mismatch vs. match), and working memory load (low load vs. high load) as within-subjects categorical factors, and IQ (z-scored) as a continuous covariate. Consistent with previous research (Baskin-Sommers, et al., 2014), there was a significant main effect of trial type, $F(1, 107) = 356.89$, $p < .001$, $\eta^2 = .77$, 95% CI: [.71, .81], indicating higher

accuracy for mismatch versus match trials. Additionally, a significant main effect of working memory load, $F(1, 107) = 128.33$, $p < .001$ , $\eta^2 = .55$, 95% CI: [.44, .62], indicated higher overall accuracy for low versus high load trials. There was also a significant two-way interaction for trial type and working memory load, $F(1,107) = 56.18$, $p < .001$, $\eta^2 = .34$, 95% CI: [.23, .44], indicating that the effect of trial type (mismatch versus match trials) was greater in the high load condition. No other task effects were significant (all *p's* > .334).

Second, the association between executive functioning and APD was examined by including APD (present vs. absent) in the GLM as a between-subjects categorical factor. There was a significant three-way reward magnitude by reward consciousness by APD interaction effect, $F(1, 106) = 4.00$, $p = .048$, $\eta^2 = .04$, 95% CI: [.00, .11]. For individuals with APD, performance was relatively better for conscious low magnitude reward trials; however, during unconscious rewards or conscious high-value rewards, individuals with APD showed relatively worse performance. Individuals without APD showed less variable performance across conditions (see also Bijleveld, Custers, & Aarts, 2011 for examples in other populations; Moors & De Houwer, 2006; Zedelius et al., 2011).

Additionally, there was a significant four-way interaction among reward magnitude, trial type, working memory load, and APD, $F(1, 106) = 5.83$, $p = .017$ , $\eta^2 = .05$, 95% CI: [.00, .13] (see Figure 4). In order to unpack this interaction, we examined the effects of the APD, reward magnitude, and working memory load on accuracy in each trial type, respectively. For match trials, there was a significant three-way interaction for APD, reward magnitude, and working memory load, $F(1, 106) = 7.30$, $p = .008$, $\eta^2 = .06$,

95% CI: [.01, .15]. Within match trials, individuals with APD were more accurate in response to high-value reward cues under low working memory load, but less accurate in response to high reward cues under high working memory load condition. By contrast, individuals without APD were less accurate in response to high reward cues in the low load condition, but more accurate in response to high reward cues in the high load condition (consistent with previous studies of healthy adults; Bijleveld et al., 2009). For mismatch trials, neither the main effect of APD nor the three-way interaction for reward magnitude, working memory load, and trial type, were significant, *p*'s > .16. Finally, neither the main effect for APD ($p = .632$) nor the five-way interaction between reward magnitude, reward consciousness, trial type, working memory load, and APD were significant ($p = .889$).

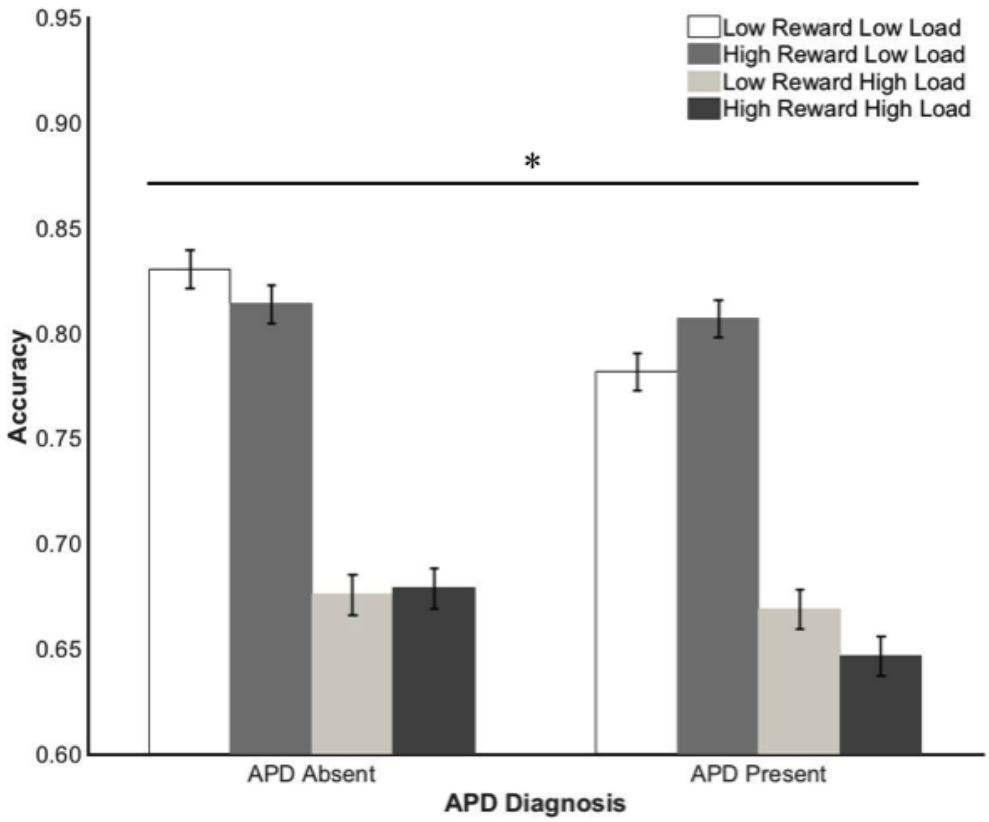


**Figure 4. Executive Functioning and APD Effects.** There was a significant four-way interaction for reward magnitude, trial type, working memory load, and APD. The effects were present in the match trials. Individuals with APD showed better performance for high versus low rewards at low load, and worse performance for high versus low rewards at high load, whereas individuals without APD showed worse performance for high versus low rewards at low load and better performance for high versus low rewards at high load. Error bars represent 1 within-subjects standard error.

## Gambling Task

First, risky choice behavior during the probabilistic decision-making task was examined in a GLM with reward magnitude (low vs. high), reward consciousness (conscious vs. unconscious), and probability (low vs. medium vs. high) as within-subjects categorical factors and IQ (z-scored) as a continuous covariate. Consistent with previous research (Hsu, Bhatt, Adolphs, Tranel, & Camerer, 2005), there was a significant main effect for reward consciousness, $F(1, 114) = 31.97$, $p < .001$, $\eta^2 = .22$, 95% CI: [.12, .32],

suggesting individuals chose risky options more often when reward information (i.e., reward magnitude) was presented

consciously. Consistent with previous research (Estle et al., 2006), there was a significant main effect for probability, $F(1.49, 169.36)$[6] = 67.75, $p < .001$, $\eta^2 = .37$, 95% CI: [.28, .45], suggesting individuals chose risky options when the probability of winning was higher, with percentage of risky choices highest on high probability, followed by medium probability and low probability trials. There also was a significant two-way interaction between reward magnitude and probability, $F(2, 228) = 5.53$, $p = .005$, $\eta^2 = .05$, 95% CI: [.01, .09], indicating a greater percentage of risky choices for low vs. high rewards at low and medium probabilities, but for high probability gambles, the risky option was chosen more often for high vs. low rewards. Lastly, the two-way interaction between probability and reward consciousness approached significance, $F(2, 228) = 3.00$, $p = .052$, $\eta^2 = .03$, 95% CI: [.00, .06], indicating a trend toward greater effects of reward consciousness when reward probability was low.

Second, the association between decision-making and APD was examined by including APD (present vs. absent) in the GLM as a between-subjects categorical factor. There was no significant main effect for APD diagnosis ($p = .925$) and no significant interactions including APD (all $p$'s > .20).

## Discussion

Previous research highlights executive dysfunction and reward hypersensitivity as core factors contributing to the behavioral dysfunction apparent in individuals with APD.

[6] Mauchly's test indicated that the assumption of sphericity had been violated for this effect, $\chi^2(2) = 48.03$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .74$).

Although these indeed are important factors to consider for APD, the present results suggest that this broad conceptualization is underspecified. Here, we identify complex interactions containing multiple factors within cognition and reward that are important for precisely understanding dysfunction in APD. Specifically, in individuals with APD, high-value rewards were disruptive during both perception and inhibition under high cognitive load. In addition, in these individuals, conscious awareness of high-value rewards was associated with reduced overall executive functioning performance. However, individuals with APD did not show abnormal probabilistic decision-making. Together, these results highlight several important patterns to consider when studying APD and the cognitive and reward abnormalities associated with the disorder.

Although perceptual processes are not often studied in APD, a growing body of literature suggests that individuals with APD actually do have difficulty detecting basic features of their environments. Individuals with APD display problems initially perceiving information, whether they are estimating the passage of time (i.e., perceiving temporal durations; Bauer, 2001) or engaging in preattentional auditory filtering (i.e., perceiving redundancy in auditory stimuli; Lijffijt et al., 2009, 2012). The present study indicates that perceptual difficulty also is apparent when anticipating high-value rewards, regardless of the conscious awareness of reward magnitude, revealing a particular maladaptive perceptual sensitivity. Dysfunction in perceptual efficiency fundamentally changes what information is seen, attended to, and, potentially acted upon. In APD, this dysfunction may precede any abnormality during executive functioning and, in some circumstances, actually lead to failures in effectively engaging adaptive behavior.

Individuals with APD display reliable dysfunction when there are demands on inhibition (Chamberlain et al., 2016; Dolan & Park, 2002; Rubio et al., 2007; Swann et al., 2009; Zeier et al., 2012) and working memory (De Brito et al., 2013; Dolan & Park, 2002). Results from the present study suggest these dysfunctions are particularly apparent in response to high-value rewards. In one context, high-value rewards disrupt inhibition during high-load at both conscious and unconscious levels. In another context, conscious awareness of high-value rewards results in poor executive functioning more broadly. It appears that individuals with APD are less able to override maladaptive response inclinations in anticipation of high-value rewards to maintain more appropriate and personally beneficial behavior.

Taken together, APD-related reward magnitude-based dysfunction in perception and executive functioning underscores a specific cognitive profile. It appears that when anticipating a high payoff, individuals with APD struggle to manage the information in their environment accurately and efficiently, resulting in maladaptive behavior (see also Results in the Supplemental Material for a comparison of performance across tasks). It may be that both the value of the reward and awareness of high-value rewards create additional cognitive load, undermining adaptive behavior for individuals with APD. Therefore, it is inaccurate to simply say that these individuals are hypersensitive to rewards or are deficient in executive functions; rather the value of the reward is an important factor undermining their ability to notice and use information in the environment.

Beyond identifying the specific factors that contribute to dysfunction in APD, the design of the present study also affords an opportunity to reveal instances of intact

cognitive and reward functioning in these individuals. During probabilistic decision-making, individuals with and without APD similarly adjust risk-taking behavior in response to reward probability, reward magnitude, and reward consciousness (Buckholtz et al., 2017; De Brito et al., 2013; Swogger et al., 2010). Moreover, during executive functioning, individuals with APD display their best inhibition while pursuing high-value rewards under low load (see Figure 4, right panel, for inhibition accuracy in the high reward/low load condition). Across experimental contexts, individuals with APD appear able to manage their reward sensitivity and engage in adaptive behavior when under markedly less pressure, as a function of generous time allotments (e.g., 4,500 ms during decision-making compared with 1,000 ms and 2,500 ms in the perception and executive function tasks, respectively; De Brito et al., 2013; Dolan & Park, 2002; Newman, 1987; Swann et al., 2009) or reduced cognitive load (e.g., 1-back inhibition, providing explicit information about outcome values during decision-making, rather than simultaneously tapping reward learning and contingency updating; De Brito et al., 2013; Dunn et al., 2006; Mazas et al., 2000). Therefore, individuals with APD do not appear to have widespread cognitive dysfunction or reward sensitivity. Leveraging knowledge about the circumstances in which individuals with APD show typical versus aberrant behavior may be important for considering why certain interventions are more effective with these individuals than others.

Several treatment approaches are used for individuals with APD. One treatment method that seems to have positive effects in APD with comorbid substance use disorders is contingency management (CM; see Brazil, van Dongen, Maes, Mars, & BaskinSommers, 2018 for review). In CM, reinforcement contingencies are assigned to

positive behaviors (e.g., abiding by the law and maintaining abstinence from drugs) to increase their frequency based on predetermined therapeutic goals (Budney, Sigmon, & Higgins, 2001). This approach may be effective because it leverages the use of explicit, unambiguous, reward contingencies for behavior, factors that are functional in individuals with APD. However, based on the present study, it is essential to be mindful of the amount of reward being offered, as rewards above a certain threshold, in certain contexts, may inadvertently disrupt adaptive behavior in APD. Beyond CM, other intervention strategies may be worth implementing among individuals with APD to bolster processes that appear deficient. Previous studies in populations with diminished inhibitory control and working memory capacities indicate that training individuals to inhibit responses to rewarding stimuli (e.g., alcohol and high-calorie foods; Houben, Havermans, Nederkoorn, & Jansen, 2012; Houben & Jansen, 2011) or maintain and update progressively larger cognitive sets in working memory (Bickel, Yi, Landes, Hill, & Baxter, 2011; Houben, Wiers, & Jansen, 2011) can lead to reductions in maladaptive behavior. Therefore, by working to remediate processes identified as suboptimal in APD, an alternative or complementary intervention strategy may be to directly target their deficits.

Several methodological and conceptual limitations should be noted. First, in an effort to study how differences in reward magnitude and consciousness affect behavior for individuals with APD, we compared responses to high versus low rewards, rather than comparing responses to rewards versus no rewards. Although our method allowed for an investigation of how individuals respond to rewards of various sizes, we cannot make conclusions about reward sensitivity among individuals with APD in the presence versus

the absence of rewards. Previous research established that APD was associated with differential responses to reward (vs. no reward) but had not specified particular dimensions of reward; therefore, we focused on reward magnitude and reward consciousness. Second, the present sample consisted of adult male offenders only, which may limit the generalizability of these findings to other populations. Future research is needed to examine specific factors of cognition and reward in other samples with APD, such as individuals who are at-risk for the disorder and female offenders. Third, it is worth considering whether the nonmonetary rewards (i.e., points and leader board rankings) used in the present study were adequate sources of reinforcement compared to real monetary rewards. Evidence suggests that points and leader boards do enhance motivation and affect psychological and behavioral outcomes (Hamari, Koivisto, & Sarsa, 2014). Nevertheless, future work should attempt to replicate the present findings using monetary rewards, while also considering ethical guidelines concerning payment for incarcerated samples. Finally, it is important to note that results from the separate tasks in the current study accounted for only a modest proportion of variance (4%–6%) in behavior. However, when estimating behavior across experimental tasks, the proportion of variance explained was slightly larger (7%). Thus, in isolation, dysfunction within specific cognitive-affective factors are unlikely to be necessary or sufficient to generate psychiatric illness (Holmes & Patrick, 2018); however, considering mechanisms as multifactorial increases the potential of more fully capturing the risk associated with specific behaviors and illness (Zalta & Shankman, 2016).

In sum, the present study indicates that complex interactions among cognitive and reward factors contribute to the behavior of individuals with APD. Hypersensitivity to

high-value rewards during perceptual and executive function efforts confer a risk factor that may contribute to chronic engagement in antisocial behaviors despite their consequences (e.g., incarceration or overdosing) in individuals with APD. Specifying the factors that account for the maladaptive behavior in APD is crucial for advancing our conceptualization of the disorder and identifying effective and targeted intervention strategies.

**Table: Study 1**

**Table 1**

*Sample characteristics and task statistics*

| | *N* | *M* | *SD* | Min | Max |
|---|---|---|---|---|---|
| Age | 116 | 34.52 | 9.75 | 20 | 58 |
| Sex (Male) | 116 | | | | |
| Race | | | | | |
| White | 52 | | | | |
| Black | 60 | | | | |
| American Indian | 1 | | | | |
| Native Hawaiian or Pacific Islander | 2 | | | | |
| Biracial | 1 | | | | |
| Ethnicity | | | | | |
| Hispanic | 20 | | | | |
| Not Hispanic | 96 | | | | |
| Highest Level of Education | | | | | |
| Grade 8 and below | 11 | | | | |
| Some High School | 62 | | | | |
| High School Diploma | 35 | | | | |
| Some College | 5 | | | | |
| College Degree | 2 | | | | |
| Graduate Degree | 1 | | | | |
| IQ | 116 | 106.11 | 9.92 | 83 | 128 |
| CD Symptom Count | 116 | 3.86 | 3.22 | 0 | 12.00 |
| Adult Antisocial Symptom Count | 116 | 3.92 | 1.61 | 0 | 7.00 |
| APD Diagnosis | | | | | |
| Absent | 58 | | | | |
| Present | 58 | | | | |
| **Visual Search Task IES by condition** | 116 | | | | |
| Unconscious Low Reward | | 0.60 | 0.07 | 0.46 | 0.87 |
| Unconscious High Reward | | 0.60 | 0.07 | 0.47 | 0.90 |
| Conscious Low Reward | | 0.59 | 0.07 | 0.44 | 0.83 |
| Conscious Reward High Reward | | 0.59 | 0.07 | 0.46 | 0.82 |
| ***n*-back Task Accuracy** | 109 | | | | |
| Match (Infrequent) Trials | | | | | |
| Low Load Unconscious Low Reward | | 0.80 | 0.18 | 0.25 | 1.00 |
| Low Load Unconscious High Reward | | 0.81 | 0.17 | 0.38 | 1.00 |
| Low Load Conscious Low Reward | | 0.81 | 0.17 | 0.29 | 1.00 |
| Low Load Conscious High Reward | | 0.82 | 0.18 | 0.25 | 1.00 |
| High Load Unconscious Low Reward | | 0.67 | 0.22 | 0.10 | 1.00 |

| | | | | | |
|---|---|---|---|---|---|
| High Load Unconscious High Reward | | 0.67 | 0.23 | 0.10 | 1.00 |
| High Load Conscious Low Reward | | 0.66 | 0.22 | 0.00 | 1.00 |
| High Load Conscious High Reward | | 0.66 | 0.22 | 0.13 | 1.00 |
| Mismatch (Frequent) Trials | | | | | |
| Low Load Unconscious Low Reward | | 0.98 | 0.03 | 0.83 | 1.00 |
| Low Load Unconscious High Reward | | 0.98 | 0.03 | 0.80 | 1.00 |
| Low Load Conscious Low Reward | | 0.99 | 0.02 | 0.90 | 1.00 |
| Low Load Conscious High Reward | | 0.99 | 0.04 | 0.73 | 1.00 |
| High Load Unconscious Low Reward | | 0.94 | 0.07 | 0.62 | 1.00 |
| High Load Unconscious High Reward | | 0.94 | 0.06 | 0.73 | 1.00 |
| High Load Conscious Low Reward | | 0.94 | 0.06 | 0.67 | 1.00 |
| High Load Conscious High Reward | | 0.94 | 0.07 | 0.70 | 1.00 |
| **Gambling Task Percent Risky Choices** | 116 | | | | |
| Low Probability Gambles | | | | | |
| Unconscious Low Reward | | 0.25 | 0.23 | 0.00 | 0.92 |
| Unconscious High Reward | | 0.22 | 0.22 | 0.00 | 1.00 |
| Conscious Low Reward | | 0.30 | 0.25 | 0.00 | 0.92 |
| Conscious Reward High Reward | | 0.30 | 0.25 | 0.00 | 0.92 |
| Medium Probability Gambles | | | | | |
| Unconscious Low Reward | | 0.32 | 0.24 | 0.00 | 1.00 |
| Unconscious High Reward | | 0.30 | 0.22 | 0.00 | 0.92 |
| Conscious Low Reward | | 0.36 | 0.25 | 0.00 | 1.00 |
| Conscious Reward High Reward | | 0.32 | 0.25 | 0.00 | 0.92 |
| High Probability Gambles | | | | | |
| Unconscious Low Reward | | 0.40 | 0.26 | 0.00 | 1.00 |
| Unconscious High Reward | | 0.42 | 0.27 | 0.00 | 1.00 |
| Conscious Low Reward | | 0.45 | 0.25 | 0.00 | 1.00 |
| Conscious Reward High Reward | | 0.49 | 0.27 | 0.00 | 1.00 |

## References: Study 1

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)* (5th, text revision ed.). Washington, D. C.: American Psychiatric Pub.

Barkataki, I., Kumari, V., Das, M., Sumich, A., Taylor, P., & Sharma, T. (2008). Neural correlates of deficient response inhibition in mentally disordered violent individuals. *Behavioral Sciences & the Law, 26*(1), 51-64.

Baskin-Sommers, A., Krusemark, E. A., Curtin, J. J., Lee, C., Vujnovich, A., & Newman, J. P. (2014). The impact of cognitive control, incentives, and working memory load on the P3 responses of externalizing prisoners. *Biological Psychology, 96*, 86-93. doi: 10.1016/j.biopsycho.2013.12.005

Bauer, L. O. (2001). Antisocial personality disorder and cocaine dependence: their effects on behavioral and electroencephalographic measures of time estimation. *Drug and alcohol dependence, 63*(1), 87-95.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to huamn prefrontal cortex. *Cognition, 50*, 7-15. doi: 0010-0277(93)00606-8

Beilock, S. L. (2007). Understanding skilled performance: Memory, attention, and 'choking under pressure'. *Sport & exercise psychology: International perspectives (pp. 153Á166). Morgantown, WV: Fitness Information Technology*.

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior, 81*(2), 179-209.

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology, 9*(1), 65-73. doi: 10.1016/j.coph.2008.12.014

Berridge, K. C., & Winkielman, P. (2003). What is an unconscious emotion?(The case for unconscious" liking"). *Cogn Emot, 17*(2), 181-211.

Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F., & Baxter, C. (2011). Remember the future: working memory training decreases delay discounting among stimulant addicts. *Biological Psychiatry, 69*(3), 260-265.

Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological Science, 20*(11), 1313-1315.

Bijleveld, E., Custers, R., & Aarts, H. (2010). Unconscious reward cues increase invested effort, but do not change speed–accuracy tradeoffs. *Cognition, 115*(2), 330-335.

Bijleveld, E., Custers, R., & Aarts, H. (2011). Once the money is in sight: Distinctive effects of conscious and unconscious rewards on task performance. *Journal of Experimental Social Psychology, 47*(4), 865-869.

Brazil, I. A., van Dongen, J. D. M., Maes, J. H. R., Mars, R. B., & Baskin-Sommers, A. (2016). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews*. doi: http://dx.doi.org/10.1016/j.neubiorev.2016.10.010

Brennan, G. M., Stuppy-Sullivan, A. M., Brazil, I. A., & Baskin-Sommers, A. R. (2017). Differentiating patterns of substance misuse by subtypes of antisocial traits in

male offenders. *The Journal of Forensic Psychiatry & Psychology, 28*(3), 341-356. doi: 10.1080/14789949.2017.1280072

Buckholtz, J. W., Karmarkar, U., Ye, S., Brennan, G. M., & Baskin-Sommers, A. R. (2017). Blunted ambiguity aversion during cost-benefit decisions in antisocial individuals. *Scientific reports, 7*.

Budney, A. J., Sigmon, S. C., & Higgins, S. T. (2001). Contingency Management *Addiction recovery tools: A practical handbook* (pp. 147).

Burgess, P. W. (1997). Theory and methodology in executive function research. *Methodology of frontal and executive function*, 81-116.

Bustin, G. M., Quoidbach, J., Hansenne, M., & Capa, R. L. (2012). Personality modulation of (un) conscious processing: Novelty Seeking and performance following supraliminal and subliminal reward cues. *Consciousness and Cognition, 21*(2), 947-952.

Capa, R. L., & Bouquet, C. A. (2018). Individual differences in reward sensitivity modulate the distinctive effects of conscious and unconscious rewards on executive performance. *Frontiers in Psychology, 9*, 148.

Capa, R. L., Bustin, G. M., Cleeremans, A., & Hansenne, M. (2011). Conscious and unconscious reward cues can affect a critical component of executive control. *Experimental psychology*.

Chamberlain, S. R., Derbyshire, K. L., Leppink, E. W., & Grant, J. E. (2016). Neurocognitive deficits associated with Antisocial Personality Disorder in non-treatment-seeking young adults. *Journal of the American Academy of Psychiatry and the Law, 44*(2), 218-225.

Congdon, E., Mumford, J. A., Cohen, J. R., Galvan, A., Canli, T., & Poldrack, R. A. (2012). Measurement and reliability of response inhibition. *Frontiers in Psychology, 3*, 37.

Crowell, T. A., Kieffer, K. M., Kugeares, S., & Vanderploeg, R. D. (2003). Executive and nonexecutive neuropsychological functioning in antisocial personality disorder. *Cognitive and Behavioral Neurology, 16*(2), 100-109.

De Brito, S. A., & Hodgins, S. (2009). Executive functions of persistent violent offenders: a critical review of the literature. *The neurobiological basis of violence: Science and rehabilitation*, 167-199.

De Brito, S. A., Viding, E., Kumari, V., Blackwood, N., & Hodgins, S. (2013). Cool and hot executive function impairments in violent offenders with antisocial personality disorder with and without psychopathy. *PLoS One, 8*(6), e65566.

Dolan, M. (2012). The neuropsychology of prefrontal function in antisocial personality disordered offenders with varying degrees of psychopathy. *Psychological Medicine, 42*(8), 1715-1725.

Dolan, M., & Park, I. (2002). The neuropsychology of antisocial personality disorder. *Psychological Medicine, 32*(3), 417-427.

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews, 30*(2), 239-271.

Gansler, D. A., Jerram, M. W., Vannorsdall, T. D., & Schretlen, D. J. (2011). Does the Iowa Gambling Task measure executive function? *Archives of Clinical Neuropsychology, 26*(8), 706-717.

Garcia-Villamisar, D., Dattilo, J., & Garcia-Martinez, M. (2017). Executive functioning in people with personality disorders. *Current opinion in psychiatry, 30*(1), 36-44.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work?--a literature review of empirical studies on gamification.* Paper presented at the 2014 47th Hawaii international conference on system sciences (HICSS).

Holmes, A. J., & Patrick, L. M. (2018). The myth of optimality in clinical neuroscience. *Trends in Cognitive Sciences*.

Houben, K., Havermans, R. C., Nederkoorn, C., & Jansen, A. (2012). Beer à No-Go: Learning to stop responding to alcohol cues reduces alcohol intake via reduced affective associations rather than increased response inhibition. *Addiction, 107*(7), 1280-1287.

Houben, K., & Jansen, A. (2011). Training inhibitory control. A recipe for resisting sweet temptations. *Appetite, 56*(2), 345-349.

Houben, K., Wiers, R. W., & Jansen, A. (2011). Getting a grip on drinking behavior: training working memory to reduce alcohol abuse. *Psychological Science, 22*(7), 968-975.

Neural systems responding to degrees of uncertainty in human decision-making., 5754, 310 Cong. Rec. 1680-1683 (2005).

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review, 17*(3), 213-233.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci, 21*(16), RC159.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *J Neurosci, 25*(19), 4806-4812. doi: 10.1523/JNEUROSCI.0642-05.2005

Kristjánsson, Á., Sigurjónsdóttir, Ó., & Driver, J. (2010). Fortune and reversals of fortune in visual search: Reward contingencies for pop-out targets affect search efficiency and target repetition effects. *Attention, Perception, & Psychophysics, 72*(5), 1229-1236.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75.

Lijffijt, M., Cox, B., Acas, M., Lane, S. D., Moeller, G. F., & Swann, A. C. (2012). Differential relationships of impulsivity or antisocial symptoms on P50, N100, or P200 auditory sensory gating in controls and antisocial personality disorder. *Journal of Psychiatric Research, 46*(6), 743-750.

Lijffijt, M., Moeller, G. F., Boutros, N. N., Burroughs, S., Steinberg, J. L., Lane, S. D., & Swann, A. C. (2009). A pilot study revealing impaired P50 gating in antisocial personality disorder. *The Journal of neuropsychiatry and clinical neurosciences, 21*(3), 328-331.

Maes, J. H., & Brazil, I. A. (2013). No clear evidence for a positive association between the interpersonal-affective aspects of psychopathy and executive functioning. *Psychiatry Research, 210*(3), 1265-1274. doi: 10.1016/j.psychres.2013.09.028

Mazas, C. A., Finn, P. R., & Steinmetz, J. E. (2000). Decision-Making Biases, Antisocial Personality, and Early-Onset Alcoholisn. *Alcoholism: Clinical and Experimental Research, 24*(7), 1034-1040.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.

Mobbs, D., Hassabis, D., Seymour, B., Marchant, J. L., Weiskopf, N., Dolan, R. J., & Frith, C. D. (2009). Choking on the money: reward-based performance decrements are associated with midbrain activity. *Psychological Science, 20*(8), 955-962.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological bulletin, 132*(2), 297.

Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review, 20*(1), 113-136.

National Institute for Health Clinical Excellence. (2009). *Antisocial personality disorder: Treatment, management and prevention*: National Institute for Health and Clinical Excellence.

Newman, J. P. (1987). Reaction to Punishment in Extraverts and Psychopaths: Implications for the Impulisve Behavior of Disinhibited Individuals. *Journal of Research in Personality, 21*, 464-480.

Ogilvie, J. M., Stewart, A. L., Chan, R. C. K., & Shum, D. H. K. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology, 49*(4), 44.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46-59.

Patrick, C. J., Durbin, C. E., & Moser, J. S. (2012). Reconceptualizing antisocial deviance in neurobehavioral terms. *Development and Psychopathology, 24*(3), 1047-1071. doi: 10.1017/s0954579412000533

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron, 59*(4), 561-567.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., & Frith, C. D. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science, 316*(5826), 904-906.

Petry, N. M. (2002). Discounting of delayed rewards in substance abusers: relationship to antisocial personality disorder. *Psychopharmacology, 162*(4), 425-432. doi: 10.1007/s00213-002-1115-1

Pochon, J., Levy, R., Fossati, P., Lehericy, S., Poline, J., Pillon, B., . . . Dubois, B. (2002). The neural system that bridges reward and cognition in humans: an fMRI study. *Proceedings of the National Academy of Sciences, 99*(8), 5669-5674.

Purves, D., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., Woldorff, M. G., & Brannon, E. M. (2008). *Cognitive Neuroscience*: Sunderland: Sinauer Associates, Inc.

Quay, H. C. (1993). The psychobiology of undersocialized aggressive conduct disorder: A theoretical perspective. *Development and Psychopathology, 5*(1-2), 165-180.

Raine, A. (2018). Antisocial Personality as a Neurodevelopmental Disorder. *Annual Review of Clinical Psychology*(0).

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current Opinion in Neurobiology, 6*(2), 228-236.

Rogers, R., Everitt, B., Baldacchino, A., Blackshaw, A., Swainson, R., Wynne, K., . . . Booker, E. (1999). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. *Neuropsychopharmacology, 20*(4), 322-339.

Rowe, D. C. (1997). Are parents to blame? A look at the antisocial personalities. *Psychological Inquiry, 8*(3), 251-260.

Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., . . . Coffey, C. E. (2002). Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. *The Journal of neuropsychiatry and clinical neurosciences, 14*(4), 377-405.

Rubio, G., Jimenez, M., Rodriguez-Jimenez, R., Martinez, I., Iribarren, M. M., Jimenez-Arriero, M. A., . . . Avila, C. (2007). Varieties of impulsivity in males with

alcohol dependence: the role of Cluster-B personality disorder. *Alcoholism: Clinical and Experimental Research, 31*(11), 1826-1832. doi: 10.1111/j.1530-0277.2007.00506.x

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*(4), 532.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol., 57*, 87-115.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*(5408), 1657-1661.

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology, 6*(328). doi: 10.3389/fpsyg.2015.00328

Stevens, M. C., Kaplan, R. F., & Hesselbrock, V. M. (2003). Executive–cognitive functioning in the development of antisocial personality disorder. *Addictive behaviors, 28*(2), 285-300.

Stuss, D. T., & Knight, R. T. (2002). *Principles of frontal lobe function*: Oxford University Press.

Swann, A. C., Lijffijt, M., Lane, S. D., Steinberg, J. L., & Moeller, F. G. (2009). Trait impulsivity and response inhibition in antisocial personality disorder. *Journal of Psychiatric Research, 43*(12), 1057-1063. doi: 10.1016/j.jpsychires.2009.03.003

Swogger, M. T., Walsh, Z., Lejuez, C. W., & Kosson, D. S. (2010). Psychopathy and Risk Taking among Jailed Inmates. *Criminal Justice and Behavior, 37*(4), 439-452. doi: 10.1177/0093854810361617

Van Gaal, S., & Lamme, V. A. (2012). Unconscious high-level information processing implication for neurobiological theories of consciousness. *The neuroscientist, 18*(3), 287-301.

Vollm, B., Richardson, P., McKie, S., Reniers, R., Elliott, R., Anderson, I. M., . . . Deakin, B. (2010). Neuronal correlates and serotonergic modulation of behavioural inhibition and reward in healthy and antisocial individuals. *Journal of Psychiatric Research, 44*(3), 123-131. doi: 10.1016/j.jpsychires.2009.07.005

Voon, V., Hassan, K., Zurowski, M., Duff-Canning, S., De Souza, M., Fox, S., . . . Miyasaki, J. (2006). Prospective prevalence of pathologic gambling and medication association in Parkinson disease. *Neurology, 66*(11), 1750-1752.

Wolfe, J. M. (1998). Visual search. *Attention, 1*, 13-73.

Zalta, A. K., & Shankman, S. A. (2016). Conducting psychopathology prevention research in the RDoC era. *Clinical Psychology: Science and Practice, 23*(1), 94-104.

Zedelius, C. M., Veling, H., & Aarts, H. (2011). Boosting or choking–How conscious and unconscious reward processing modulate the active maintenance of goal-relevant information. *Consciousness and Cognition, 20*(2), 355-362.

Zedelius, C. M., Veling, H., Custers, R., Bijleveld, E., Chiew, K. S., & Aarts, H. (2014). A new perspective on human reward research: How consciously and

unconsciously perceived reward information influences performance. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 493-508.

Zeier, J. D., Baskin-Sommers, A., Hiatt Racer, K. D., & Newman, J. P. (2012). Cognitive control deficits associated with antisocial personality disorder and psychopathy. *Personality Disorders: Theory, Research, and Treatment, 3*(3), 283-293. doi: 10.1037/a0023137

# Chapter 3: Study 2

## Aberrant cost-benefit integration during effort-based decision-making relates to antisocial personality disorder symptoms

## Abstract

Aberrant cost-benefit decision-making is often implicated in antisocial behavior. Previous research highlights how delay and ambiguity sensitivity are associated with chronic antisocial behavior; however, other forms of cost-benefit decision-making—effort-based choice—have received less attention. We administered the Effort Expenditure for Rewards Task in a community sample enriched for antisocial behavior (*N*=80). Individuals with more antisocial personality disorder symptoms were less likely to use information about expected value when deciding between high effort/high reward and low effort/low reward options. Additionally, their behavior was better explained by a simple computational model of effort-based decision making that did not incorporate information about reward and probability compared to more complex models that integrated this information. Further, individuals with more antisocial personality disorder symptoms who were experiencing high levels of negative affect during the experimental session and individuals with heightened sensitivity to delay costs during inter-temporal decision-making were the least sensitive to expected value signals when making decisions to engage in effortful behavior. Together, these findings suggest that antisociality is related to difficulty integrating multiple decision variables to guide behavior during effort-based decision-making.

Individuals with antisocial personality disorder (APD) engage in a lifelong pattern of impulsive and antisocial behavior. Their irresponsible, aggressive, and reckless actions frequently result in significant consequences—from poor physical and mental health outcomes (Black, Gunter, Loveless, Allen, & Sieleni, 2010; Goldstein et al., 2008) to high rates of unemployment and incarceration (Knapp, King, Healey, & Thomas, 2011). Moreover, individuals with APD are resistant to treatment and prone to recidivism (Raine, 2018). Their apparent inability to weigh the costs of continuing antisocial acts against the potential benefits of desisting suggests that they suffer from fundamental decision-making abnormalities (Gregory et al., 2015). Indeed, empirical evidence indicates that impaired cost-benefit decision-making may be a key factor influencing behavior among antisocial individuals (Buckholtz, 2015).

Broadly speaking, one can differentiate distinct facets of cost-benefit decision-making according to the specific costs that an individual must integrate in order to make optimal decisions (e.g. risk, delay, ambiguity; effort; Rudebeck, Walton, Smyth, Bannerman, & Rushworth, 2006). More specifically related to antisocial behavior, abnormal decision making in the presence of delay and ambiguity costs are documented among youth with Conduct Disorder (CD; a precursor to APD) and adults with APD (Buckholtz, Karmarkar, Ye, Brennan, & Baskin-Sommers, 2017; Petry, 2002; White et al., 2014). In youth and adulthood, individuals with APD are hypersensitive to delays in reward receipt and consequently choose smaller immediate reward options over larger delayed reward options when making decisions (Petry, 2002; White et al., 2014). They also are relatively insensitive to ambiguity surrounding choice options and are less

deterred by options where outcome probabilities are unknown or undiscoverable (Buckholtz et al., 2017; Hobson, Scott, & Rubia, 2011; Mazas, Finn, & Steinmetz, 2000).

While previous research sheds important light on how delay and ambiguity sensitivity relate to antisocial behavior, other forms of cost-benefit decision-making – especially those involving effort-based choice – have received less attention. Effort-based decision-making describes the process of choosing how much effort to invest in order to obtain a valued outcome and may involve choosing between options with varying work requirements (Chong, Bonnelle, & Husain, 2016; Salamone, Correa, Farrar, & Mingote, 2007). Within the framework of cost-benefit decision-making, effort can be considered a cost (c.f., Inzlicht, Shenhav, & Olivola, 2018). Consistent with the notion that adjusting effort expenditure based on expected value is an essential, evolutionarily conserved, component of adaptive choice behavior (Salamone et al., 2007), effort-cost discounting of subjective value has been demonstrated across multiple species (Chong et al., 2016; Kurniawan, Guitart-Masip, & Dolan, 2011; Phillips, Walton, & Jhou, 2007; Walton, Rudebeck, Bannerman, & Rushworth, 2007).

Recent work using effort-based decision-making tasks points to aberrant effort-based computations as a proximal cognitive mechanism underlying motivation-related symptoms in diverse forms of clinical disorders (Salamone et al., 2016; Treadway, Bossaller, Shelton, & Zald, 2012a; Treadway & Zald, 2013). For example, individuals with schizophrenia demonstrate a relative insensitivity to information about reward magnitude and probability during effort-based decision-making. Moreover, this insensitivity is associated with more severe negative symptoms and functional outcomes, suggesting effort-based computations may be related to variability in the expression of

schizophrenia (see Culbreth, Moran, & Barch, 2018, for review). Additionally, patients with major depressive disorder show similar deficits in effort-based decision-making, with the magnitude of these deficits tracking severity of anhedonia and duration of depressive episodes (Treadway et al., 2012a). Together, this work suggests that effort-based decision-making deficits may represent a transdiagnostic factor important for the expression and course of clinical disorders. However, despite evidence for cost-benefit decision-making deficits in those who chronically engage in antisocial behavior, effort-based decision-making remains relatively unexplored.

The goal of the present study was to examine the association between effort-based decision-making and antisociality. Separate lines of research suggest that individuals with disorders associated with antisociality (e.g., borderline personality disorder, Daughters, Sargeant, Bornovalova, Gratz, & Lejuez, 2008; substance use disorders, Krueger, Markon, Patrick, & Iacono, 2005) have difficulty persisting in rewarded tasks when physical or cognitive effort costs are high (Bornovalova, Gratz, Daughters, Hunt, & Lejuez, 2012; Brandon et al., 2003; Goldstein & Volkow, 2011). Adding to the likelihood that effort-based decision-making is compromised in antisocial individuals, there is a striking overlap between the neural circuitry involved in effort-based decision-making (e.g., the anterior cingulate cortex; ACC; Croxson, Walton, O'Reilly, Behrens, & Rushworth, 2009; the dorsolateral prefrontal cortex; DLPFC; Goldstein & Volkow, 2011; the mesolimbic dopamine system; Treadway et al., 2012b) and circuit-level abnormalities associated with chronic antisociality (i.e., DA system dysfunction; Buckholtz et al., 2010; Ponce et al., 2003; reductions in ACC and DLPFC gray matter and activity; Raine, Buchsbaum, & LaCasse, 1997; Rosenbloom, Schmahmann, & Price, 2012; Yang &

Raine, 2009). Together, this research provides a premise for the hypothesis that the integration of cost and benefit signals during effort-based decision-making may be disrupted in chronically antisocial individuals.

To investigate the association between antisociality and effort-based decision-making, we administered the Effort Expenditure for Rewards Task (EEfRT) in a community sample enriched for antisocial behavior. We focused on examining the number of antisocial personality disorder symptoms, based on previous work in other clinical populations linking variability in cumulative symptom expression and dysfunction in effort-based decision-making (e.g., Culbreth et al., 2018; Treadway et al., 2012a; Yang et al., 2014). We explored the relationship between antisociality and effort-based choice using multiple modeling techniques to examine how individuals with higher levels of antisociality incorporate cost-benefit information when choosing whether to exert effort. In addition, we used a self-report measure of positive and negative affect to determine the relevance of state-level variation in affect for antisocial individuals during effort-based decision-making, based on the link between negative affect and antisociality, as well as previous studies indicating effects of negative affect on effort-based decision-making. Finally, given the known importance of delay-cost sensitivity for antisocial individuals and the overlap between delay and effort-based decision-making, we employed a self-report measure of delay discounting to identify potential moderating effects of this related and well-documented decision-making facet.

## Method

### Participants

Participants consisted of 92 adults ages 18 to 75 recruited from the community through flyers soliciting risk-taking (e.g., substance use, crime, gambling, impulsive behavior, bullying) individuals in New Haven County, Connecticut (see Table 1). A prescreen phone interview and in-person assessment materials were used to exclude individuals who: were younger than 18 or over 75, had performed below the fourth-grade level on a standardized measure of reading (Wide Range Achievement Test-III; Wilkinson, 1993), scored below 70 on a brief measure of IQ (Zachary, 1986), had diagnoses of schizophrenia, bipolar disorder, or psychosis, not otherwise specified (Structured Clinical Interview for DSM-5 Disorders; First, Williams, Karg, & Spitzer, 2015), or had a history of certain medical problems (e.g., uncorrectable auditory or visual deficits; head injury with loss of consciousness greater than 30 minutes) that may impact their comprehension of the materials or performance on the task. All participants provided written informed consent according to the procedures set forth by the Yale University Human Investigation Committee. Participants earned $10/hour for their completion of the self-report measures and the experimental task. Participants also were eligible to earn a cash bonus (range $2-$8) based on the sum of two randomly selected trials from the experimental task, rounded to the nearest dollar.

We conducted an *a priori* power analysis based on previous studies of individual differences in effort-based decision-making (Treadway et al., 2012a; Treadway, Buckholtz, Schwartzman, Lambert, & Zald, 2009) using GLIMMPSE Statistical Software (Kreidler et al., 2013). The power analysis indicated that a sample size of 48-80 participants would result in the desired (80%) power to detect an effect size comparable to other studies (i.e., a 15-20% difference in hard task choices; Barch, Treadway, &

Schoen, 2014; Treadway et al., 2012a; Treadway et al., 2009) finding three-way interactions among key task variables (e.g., EV as a within-subjects repeated measure) and individual difference measures (e.g., antisocial personality disorder symptoms and moderator variables as between-subjects variables) in the EEfRT, controlling for covariates.

## Measures

**Structured Clinical Interview for DSM-5 Disorders** (SCID-5; First et al., 2015). The SCID-5 was used to determine whether individuals met criteria for the symptoms of antisocial personality disorder (APD). Including conduct disorder (CD) and adult antisocial symptoms, there are a total of 22 symptoms for the APD diagnosis. Total APD symptom scores were formed by counting the number of items coded as "3 - threshold" across CD and APD symptom criteria. The distribution of APD symptom scores was positively skewed; as a result, the APD symptom score was normalized by adding 1 and naturally log transforming the resulting sum.

**Positive-Negative Affect Scale** (PANAS; Watson, Clark, & Tellegen, 1988). Previous studies of effort-based decision-making in clinical samples indicate that negative affect is associated with reduced effortful choices (e.g., Treadway et al., 2009). Additionally, many models of antisociality highlight negative emotionality as a core feature (Raine, 2018), and poor persistence and self-regulation under negative affect are prominent risk factors for antisocial behavior (Daughters et al., 2008; Deater-Deckard, Petrill, & Thompson, 2007). To examine the role of negative affect and its moderating role in effort-based choice among antisocial individuals, we administered the PANAS, a self-report questionnaire, to tap both positive and negative affect at the time immediately

preceding the behavioral task. Participants were asked to rate the extent to which they feel each item on a 5-point Likert scale (1 = very slightly or not at all, 2 = a little, 3 = moderately, 4 = quite a bit, 5 = extremely). The scores on the ten positive affect items (e.g., interested, excited, strong) were added, with higher scores indicating higher positive affect. The scores on the ten negative items (e.g., distressed, upset, nervous) were added, with higher scores indicating higher negative affect. Positive and negative affect scores were normalized using natural-log transformations.

**Monetary Choice Questionnaire** (MCQ; Kirby, Petry, & Bickel, 1999). Chronically antisocial youth and adults show elevated delay discounting behavior during inter-temporal choice (Petry, 2002; White et al., 2014). Previous studies identify both shared and cost-selective neurobiological mechanisms underlying delay discounting and effort-based decision-making (Peters & Buchel, 2011; Prévost, Pessiglione, Météreau, Cléry-Melin, & Dreher, 2010), raising the possibility that variation in delay-cost sensitivity might influence effort-based computations. Of note, decisions about effort allocation often involve an implicit consideration of delay. Tasks that require more effort typically take longer to complete, and therefore, provide delayed payouts. Thus, we used the MCQ, a 27-item questionnaire, to measure delay discounting.

For each binary choice item, participants indicated their preference between a larger amount of money ($25-$85) available at a delay (7–186 days) and a smaller amount of money ($11-$80) available immediately. Discount rates, *k*, were calculated according to a hyperbolic discounting function (Mazur, 1987), $V_d = r / (1 + kD)$, where $V_d$ was the subjective value of a delayed reward of magnitude *r* available at delay *D*. Distributions of *k* estimations were positively skewed and thus were natural log-

transformed. Higher ln*k* values reflect a greater tendency to value immediate rewards over delayed rewards.

**Effort Expenditure for Rewards Task** (EEfRT; Treadway et al., 2009). The EEfRT, a multi-trial computerized button-pressing game, measured the extent to which individuals were willing to incur greater effort costs in order to obtain larger, more probable rewards (see for examples Barch et al., 2014; Treadway et al., 2012a; Treadway et al., 2012b). Participants made a series of choices between completing an easy task and a hard task for variable amounts of reward. The hard task always required 100 button presses to be made within 30s, with the non-dominant pinkie finger. The easy task always required 30 button presses to be made within 7s, with the dominant index finger. Each trial started with an information/decision screen indicating the reward magnitude that could be earned for completing the easy task (always $1.00) and hard task (from $1.24 - $4.30), as well as the probability that completing either task would result in earning the reward (either a 12%, 50%, or 88% probability of being rewarded; see Figure 1). The participant had five seconds to make a choice between the easy and hard task for each trial. If the participant did not make a choice within five seconds, the computer randomly selected a choice for the participant. After the choice period, a button press screen appeared, and the participant completed button presses for the selected task. Individuals received feedback about whether they successfully completed the selected task, and whether they earned a reward on each trial. Participants completed four practice trials prior to the task beginning and were monitored by research assistants via video camera to ensure proper execution of button presses and engagement with the EEfRT.

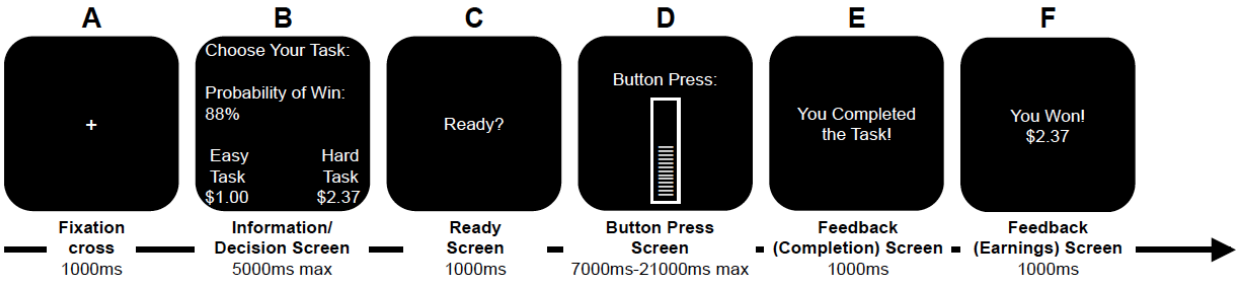


**Figure 2. Sample of trial sequence for the EEfRT.** Each trial began with (**A**) a fixation cross (1000ms) followed by (**B)** a screen on which information about the probability of earning a reward and the reward magnitude for completing each task was presented (up to 5000ms), and individuals were asked to make a decision via button press. After either a choice was made or the allotted decision time was up, (**C)** a ready screen was presented briefly (1000ms), followed by (**D)** a button press screen, which marked each button press the participant made during the allotted task time (up to 7000ms or 21000ms). After the task was completed or the allotted task time had passed, participants saw (**E)** a feedback screen indicating whether they successfully completed the task (1000ms) and, if they completed the task, (**F)** a second feedback screen indicating the reward amount, if any, they earned for completing the task (1000ms).

Following previous research (e.g., Treadway et al., 2012a; Treadway et al., 2009), trial-by-trial modeling was conducted to account for time-varying parameters, only the first 50 trials after practice were extracted for data analysis, and only trials where the participant (not the computer) made a choice were included in analyses**.** Moreover, participants were excluded completely if they had a physical feature that precluded complete engagement with the task (e.g., broken finger, arthritis in wrist), technical problems during their session (e.g., computer crash in the middle of the task), or behavior that indicated they were not making decisions or were not completing the selected tasks (i.e., timing out on over 10% of trials, failing to complete the selected task on over 20% of trials). The final sample consisted of 80 participants. Excluded participants did not differ significantly from included participants on the number of antisocial personality disorder symptoms ($p$ = .306).

The primary outcome from the EEfRT was choice (easy task versus hard task). For each trial, we incorporated two key variables into analyses: the reward magnitude at stake for the hard task and the probability of earning a reward for successfully completing

either task. Additionally, we calculated the product of these two variables to represent the expected value (EV) of the hard task for each trial. This third key variable, EV, thus represented the combined impact of reward magnitude and probability.

### *Computational Modeling*

In addition to mixed-effects logistic regression models including EV as a key variable, we used a newly developed computational modeling approach to quantify the extent to which individuals used available information about reward and probability when deciding to allocate effort (Cooper et al., 2019). Using this approach, we fit and compared three models that reflected different strategies for allocating effort to each individual's choice behavior on the EEfRT.

#### *Full Subjective Value (SV) model*

The first model, a full subjective value (SV) model, assumes participants incorporate both reward magnitude and probability information when making decisions to invest effort. The full SV model fits best for participants whose choices to allocate effort are most consistently influenced by trial-wise reward and probability information. According to the full SV model, the SV of a given trial is calculated by reducing the objective reward, R ($1 to $4.30), by the probability of obtaining it, P, and the amount of effort, E, required to obtain it (.3 for easy trials, 1 for hard trials; Eq 1). Individual differences in the extent to which the reward is discounted by probability and effort are captured by free parameters that weigh each of the components.

$$SV = R*P^{h} - kE \qquad \text{Eq 1}$$

Higher values of *k* reflect greater effort aversion (i.e., perceiving effort to be very costly). Higher values of *h* reflect greater risk aversion, with greater weighting of probability on subjective value.

Using the Softmax decision rule (Sutton & Barto, 1998), subjective values are transformed into probabilities of selecting each option. Here, *t* is an inverse temperature parameter that reflects a tendency to favor options with higher subjective values (Eq 2):

$$\text{p(hard)} = \frac{e^{\text{SVhard}\cdot t}}{e^{\text{SVhard}\cdot t} + e^{\text{SVeasy}\cdot t}}$$

Eq 2

Accordingly, the full SV model has three free parameters: *k*, *h*, and *t*. The *k* parameter decreases subjective value based on the amount of required effort, the *h* parameter reduces subjective value according to the probability of obtaining the reward, and the *t* parameter influences the extent to which choice behavior coincided with options with higher subjective values.

Following procedures described by Cooper and colleagues (2019), a fit of the full SV model with *h* constrained to 1 also was fit to account for individuals who integrate reward, effort, and probability without distorting probability (i.e., Eq 1 where free parameter *h* is held constant at 1). This practice prevented overpenalizing model fit for the additional free parameter *h*. Participants best fit by the SV model with either a flexible *h* parameter or with *h* set to 1 were all included in the full SV model group, since fit to either model represents the integration of reward, effort, and probability during decision making.

*Reward-Only Model*

A reward-only SV model assumes that participants only incorporate trial-wise

changes in reward-magnitude when allocating effort, since previous studies using the EEfRT indicated that some participants allocate effort based only on rewards, ignoring probability information (Cooper et al., 2019). For these participants, a simpler model that does not incorporate a free parameter for scaling probability information captures choice behavior more accurately. The reward-only SV model essentially represents the full SV model when *h* assumes a value of zero (Eq 3).

$$SV = R - kE \quad \text{Eq 3}$$

Although both the reward-only SV model and the full SV model with $h = 1$ hold *h* constant and have the same number of free parameters, they are interpreted very differently. Whereas restricting *h* to 0 represents choice behavior that is unaffected by probability information, *holding h* = 1 allows for probability information to influence subjective value.

*Bias Model*

A bias model assumes that participants do not consider reward or probability information when allocating effort. This model is the least complex model, containing only 1 free parameter, *b*, which represents a bias towards the low-effort option. The probability of selecting the high-effort option is simply 1- *b*. This model assumes a consistent probability of choosing the low-effort option across trials, regardless of probability or reward. This model provides a similar or better fit than the SV model for participants who highly favor one option, respond randomly, or whose choice behavior is inconsistent with the assumptions of the SV model (i.e. favoring effort allocation for low reward).

*Model Fitting*

Comparing the fit of these three computational models describes the extent to which participants systematically allocate effort based on all available information (full SV model), allocate effort primarily based on reward (reward only SV model), and make choices that are not strongly or consistently influenced by trial-specific information (bias model). All models were fit in MATLAB using maximum likelihood estimation with optimization function fminsearch (MATLAB and Statistics Toolbox, 2016b). Models were fit individually to each participants' data. Parameters selected for each participant optimized the likelihood of the behavioral data. For the subjective value models, *k* and *h* parameters were constrained to be between 0 and 10, and *t* was constrained between 0 and 100. All three models were fit with 1000 random parameter initializations. The three models varied in terms of their flexibility, since the SV models benefitted from the flexibility of additional free parameters. To account for differences in flexibility, we used Bayesian information criterion (BIC; Schwarz, 1978) to compare model fit for each participant. BIC penalizes models that have additional flexibility (more free parameters), and favors simpler models when log-likelihood is the same or similar. BIC incorporates goodness of fit (likelihood, L*i*), number of free parameters (*Vi*), and the number of observations (i.e., number of trials, *n;* Eq 4):

$$\mathrm{BIC}_i = -2\ln(L_i) + V_i \ln(\mathrm{n})$$

Eq 4

After calculating BIC for each of the models for each participant, we calculated BIC difference measures (ΔBIC) to quantify the improvement in goodness of fit that either the full SV model or the reward-only SV model provided over the bias model for each participant (Eq 5).

$$\Delta BIC = BIC_{BIAS} - BIC_{SV}$$

Eq 5

A positive ΔBIC indicated that a participant was better fit by the respective SV model, and that their choice behavior was better explained by incorporating trial-by-trial variability in reward (and probability, for the Full SV model). A negative ΔBIC indicated that a participant's behavior was better explained by the simpler model.

## Results

### Reward Magnitude, probability, and expected value (EV) influenced effort-based choice

We ran a mixed-effects logistic regression model in STATA 14 (StataCorp, 2015) to confirm that participants' choices to select high vs. low effort options were guided by EV. Choice was considered as a binary outcome variable (0 [Easy Task], 1 [Hard Task]), and EV, participant age, and trial number as continuous fixed-effect predictors[7]. Participant was treated as a random effect. Consistent with previous research, there was a significant main effect for EV (B = 0.801, SE = 0.041, 95% CI = 0.720, 0.882, $z$ = 19.44, $p < 0.001$) on choice behavior, such that as the EV for the hard task increased, there was a greater likelihood of selecting the hard task. Moreover, EV predicted choice behavior (B = 0.484, SE = 0.144, 95% CI = 0.203, 0.765, z = 3.37, $p$ = 0.001) even after controlling for trial-wise variation in both reward magnitude (mean-centered; B = 0.562,

[7] Consistent with prior studies using the EEfRT, age and trial number were included in all mixed model analyses to account for potential confounding effects of fatigue, including those associated with age (Treadway et al., 2012a; 2012b; Treadway et al., 2009; Wardle et al., 2011). For all results, age was a significant predictor of fewer hard task choices (all $p$s < .01) and trial number was either a significant predictor of fewer hard choices or predicted fewer hard choices at trend levels (all $p$s < .07). However, all effects reported in the manuscript remained significant when age and trial number were not included as covariates.

SE = 0.045, 95% CI = 0.474, 0.51, z = 12.45, $p < 0.001$) and probability (mean-centered; B = 2.04, SE = 0.128, 95% CI = 1.787, 2.287, z = 15.97, $p < 0.001$). These results confirmed that decisions about effort expenditure relied on the integration of multiple decision variables available at the time of choice to guide action value estimation and drive action selection.

**Individuals with more APD symptoms showed diminished cost-benefit integration during effort-based choice**

We used a mixed-effect logistic regression model and included EV, APD symptoms, and the EV-by-APD symptoms interaction as continuous fixed-effect predictors to determine whether the use of EV to guide choice behavior varied by antisociality. Consistent with our hypothesis, individuals with more APD symptoms appeared less likely to use EV to modulate effort expenditure (B = −0.174, SE = 0.043, 95% CI = -0.259, -0.089, $z = -4.02$, $p < 0.001$; Figure 2).

This pattern of behavior was not necessarily evidence of a deficit in integrating effort costs and EV. It was possible that antisociality-linked differences in the use of EV information during effort-based choice could reflect blunted sensitivity to one or both of the decision variables used to calculate EV. It was, thus, important to identify any significant two-way interactions between our measure of antisociality and reward magnitude or probability. We examined reward magnitude, reward probability, and their interactions with APD symptoms as continuous fixed effect predictors in a mixed-effect logistic regression model. Indeed, sensitivity to both reward magnitude and probability was blunted in individuals with more APD symptoms: reward magnitude and probability

showed independent interactions with APD symptoms (reward magnitude-by-APD symptoms: B = -0.199, SE = 0.048, 95% C.I. = -0.292, -0.106, z = -4.18, $p < 0.001$; probability-by-APD symptoms: B = -0.404, SE = 0.135, 95% C.I. = -0.668, -0.140, z = -3.00, $p = 0.003$).

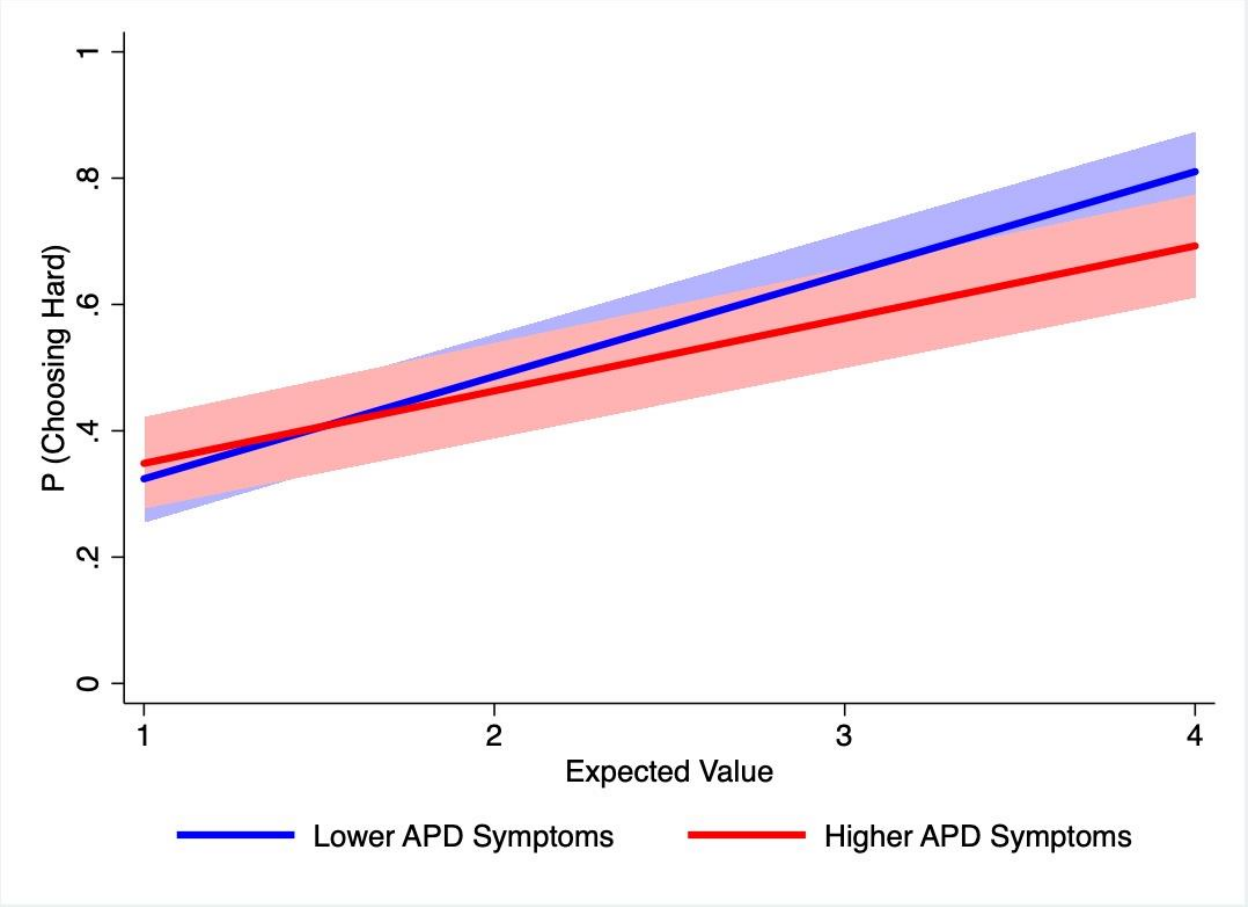


**Figure 3. Expected value (EV) by level of APD symptoms.** Individuals with more APD symptoms were less likely to use EV to modulate effort expenditure. Lines represent ±1 *SD* from the mean. Shading around lines represents 95% confidence intervals for point estimates.

To confirm that the EV-by-APD symptoms interaction (reflecting the integration of two decision variables) accounted for variance in choice behavior over and above that which could be explained by the interactions between APD symptoms and the two "simple" decision variables (reward magnitude and probability), we included the three-way interaction between reward, probability, and APD symptoms as a continuous fixed-

effect predictor. If the EV-by-APD symptom interaction truly signified an integration deficit, rather than simply reflecting an insensitivity to reward magnitude and/or probability that carried over into the EV term, the three-way interaction between reward, probability, and APD symptoms should predict choice behavior even when two-way interactions between the 'simple' decision variables and antisociality were modeled. A significant three-way interaction between reward magnitude, probability, and APD symptoms was observed (B = 0.273, SE = 0.112, 95% C.I. = 0.054, 0.493, z = 2.44, $p$ = 0.015). This pattern of results was consistent with the notion that greater levels of antisociality are associated with a relative deficit in the capacity to integrate available decision variables – here, reward magnitude and probability of reward receipt – to modulate action valuation and selection during cost-benefit decisions involving effort allocation. With initial evidence of an integration deficit beyond the independent effects of antisociality on the use of reward magnitude and probability information, we examined whether highly antisocial individuals also would demonstrate integration deficits based on computational models of effort-based decision-making.

**Individuals with more APD symptoms used simpler decisional models of effort compared to complex models to guide behavior**

To examine whether antisociality was associated with the extent to which more complex versus more simple computational models account for choice behavior, we calculated partial correlations between APD symptoms and ΔBIC scores comparing fit for both the full SV model and the reward-only model to the bias model, controlling for age. Partial correlations indicated that individuals with more APD symptoms displayed

behavior that was better explained by the simple bias model over the full SV model, $r(77) = -.23$, $p = 0.040$, and behavior that was better explained by the simple bias model over the reward-only SV model, $r(77) = -.28$, $p = 0.012$. These results are consistent with findings from the mixed-effects logistic regression models; greater levels of antisociality were associated with less use of available information during effort-based decision-making, resulting in choice behavior that was best characterized by a model that does not include reward or probability information. Having confirmed that APD symptoms were associated with reduced integration of reward and probability information using multiple modeling techniques, we used EV in subsequent mixed-effects logistic regression models to analyze the impact of negative affect and delay discounting on integration deficits during effort-based choice.

**Negative affect predicted effort and blunted EV sensitivity during effort-based decision-making among individuals with more APD symptoms**

To explore the possibility that negative affect moderated the relationship between antisociality and effort-based choice, we first examined the separate relationships among positive and negative affect and effort-based decision-making using separate mixed-effects logistic regression models. We again considered choice as a binary outcome variable (0 [Easy Task], 1 [Hard Task]), and included EV, participant age, trial number, and affect (positive or negative) as continuous fixed-effect predictors. We found no relationship between positive affect and effort-based choice ($p = 0.631$). However, we found a significant relationship between negative affect scores and the probability of making high effort/high reward choices (B = -0.896, SE = -0.452, 95% C.I. = -1.782, -

0.009, z = -1.98, $p$ = 0.048). This result was consistent with previous studies (e.g., Treadway et al., 2009) indicating that individuals with higher levels of negative affect exhibit increased effort aversion during decision-making.

We then examined whether negative affect (NA) impacted the use of EV signals during effort-based choice. The EV-by-NA interaction indicated a statistical trend (B = -0.255, SE = 0.140, 95% C.I. = -0.530, 0.19, z = -1.82, $p$ = 0.068), such that greater levels of negative affect were associated with diminished EV use during effort-based decision-making. Next, we sought to determine whether being in a negative affective state modulated the relationship between EV sensitivity and antisociality. This model included all first and second-order terms, and the three-way interaction term for negative affect, EV, and APD symptoms. The EV-by-APD symptoms-by-NA interaction was significant (B = -0.281, SE = 0.141, 95% C.I. = -.558, -0.004, z = -1.99, $p$ = 0.047; see Figure 3). This result indicated that EV sensitivity in the highest EV trials was weaker in individuals with more APD symptoms who were experiencing greater levels of negative affect at the time of the decision-making task. By contrast, individuals with more APD symptoms who were not experiencing marked negative affect showed EV integration similar to that of individuals with fewer APD symptoms during the trials with the highest EV.

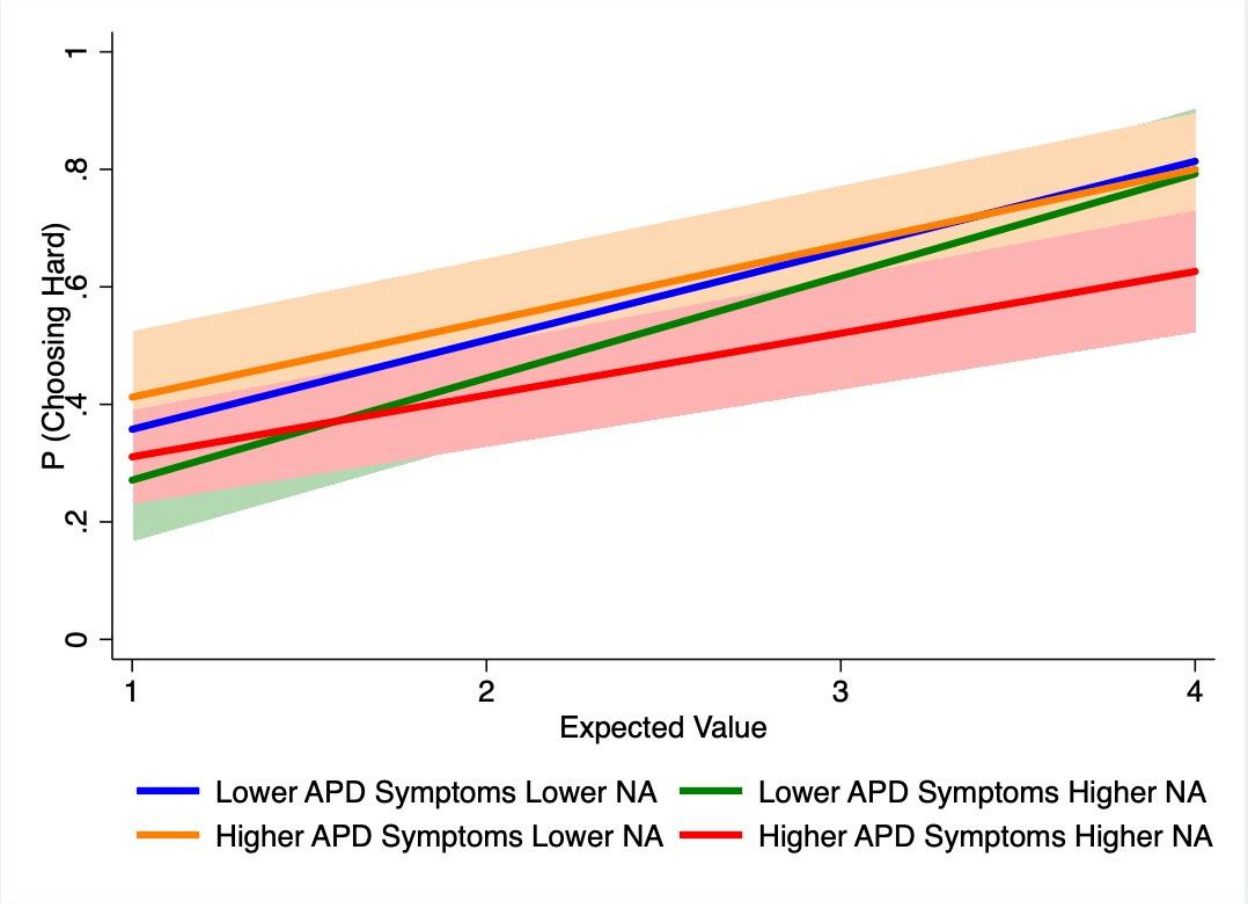


**Figure 4. Expected Value (EV) by APD symptoms and negative affect (NA).** During high EV trials, individuals with more APD symptoms experiencing higher levels of NA were less likely to use EV information during effort-based computations. Lines represent 1 SD above and below the mean for APD symptoms and NA scores. Shading around lines represents 95% CI for point estimates.

### Delay-cost sensitivity predicted effort and blunted EV sensitivity during effort-based decision-making, but did not interact with APD symptoms

We also explored the possibility that individuals who were sensitive to delay costs would be sensitive to effort costs, and that this tendency might moderate the relationship between antisociality and effort-based choice. In another mixed-effect logistic regression model, we included ln*k* as a continuous fixed effect predictor of choice behavior. We found a negative relationship between ln*k* values calculated from the MCQ and the probability of making high effort/high reward choices (B = -0.225, SE = -0.099, 95% C.I.

= -0.419, -0.031, z = -2.28, $p$ = 0.023). This result indicated that participants with higher levels of delay discounting exhibited increased effort aversion during decision-making. We next examined whether the use of EV signals to make effort-based choices varied as a function of delay discounting. A significant EV-by-ln*k* interaction was observed (B = -0.153, SE = 0.011, 95% C.I. = -0.221, -0.085, z = -4.42, $p$ < 0.001), such that EV use during effort-based decision-making diminished with greater levels of delay discounting.

Finally, modeling all first and second-order terms and the three-way interaction term among EV, ln*k*, and APD symptoms, we did not observe significant interactions between delay cost sensitivity and antisociality on effort allocation preferences (APD symptoms-by-ln*k*: B = -0.046, SE = 0.145, 95% C.I. = -0.330, 0.239, z = -0.32, $p$ = 0.752). The three-way interaction between EV, APD symptoms, and ln*k* likewise was not significant (EV-by-APD symptoms-by-ln*k*: B = 0.053, SE = 0.038, 95% C.I. = -0.021, 0.126, z = 1.41, $p$ = 0.158). On the whole, this pattern of results indicated that delay discounting impacts effort-based computations by interfering with the use of expected value information during effort-based decision-making. However, evidence supporting the relevance of this relationship for antisociality was not compelling given that we did not observe an interaction among EV, APD symptoms, and delay discounting.

## Discussion

In the present study, we found evidence of aberrant cost-benefit computations among individuals with increased antisociality. Specifically, individuals with more APD symptoms displayed decreased sensitivity to expected value information during effort-based decision-making. Of note, the significant three-way interaction among reward magnitude, probability, and antisociality reflected deficits in the integration of multiple

decision variables during effort-based decision-making. Results from the computational modeling estimates similarly showed that with increased levels of antisociality, individuals displayed behavior that was better described by a simple computational model that did not include information about reward and probability during effort-based choice. Moreover, antisociality-linked integration deficits had affective specificity; in particular, the individuals with increased levels of antisociality who were experiencing negative affect exhibited the most pronounced cost-benefit decision-making integration deficits during effort-based choice, opting to choose the easy task when the hard task had its highest expected value. These results provide evidence that the capacity to integrate information from multiple decision-making variables to estimate expected value and guide action selection is compromised in individuals with more severe antisocial behavior, and this pattern may be mood-dependent.

Our findings are especially interesting in light of the brain circuit mechanisms underlying cost-benefit decision-making, broadly, and effort-based decision-making in particular. Dopaminergic dysfunction is associated with a preference for low effort-low cost choices during effort-based decision-making (Treadway et al., 2012b) and also is reported in antisociality among youth (Matthys, Vanderschuren, & Schutter, 2013) and adults (Buckholtz et al., 2010; Ponce et al., 2003). Specifically, greater dopaminergic dysfunction is correlated with more CD symptoms, earlier onset of antisocial behaviors, greater levels of aggression, and a greater number of criminal offenses (Caspi et al., 2008; Matthys et al., 2013; Thapar et al., 2005). Dopaminergic dysfunction also is associated with variability in the expression of other clinical disorders (e.g., Culbreth et al., 2018; Treadway et al., 2012a; Yang et al., 2014), including substance use disorders

(Martinez et al., 2011), which are highly comorbid and genetically linked with antisociality (Krueger et al., 2005). Accordingly, our finding that individuals with more APD behaviors show deficits in decision variable integration during effort-based choice is consistent with the possibility that aberrant decision-making preferences may play a role in the onset, maintenance, and modifiability of more severe antisociality.

Furthermore, the disruption of decision variable integration among individuals with increased levels of antisociality appears to be mood-dependent, highlighting the importance of considering negative affect as a determinant of the modulation of effort expenditure. Research in rodents and humans documents that negative affect, from acute stress to clinical levels of depression, impairs effort-based decision-making (Shafiei, Gray, Viau, & Floresco, 2012; Treadway et al., 2012a). The relationship between negative affect and antisociality also is well documented, with considerable evidence supporting a positive association between negative affect, chronic distress, and chronic antisocial behavior (Hyde, Byrd, Votruba-Drzal, Hariri, & Manuck, 2014; Lorber, 2004), including aggression and delinquency in youth (Daughters et al., 2008; Deater-Deckard et al., 2007; Sontag, Graber, Brooks-Gunn, & Warren, 2008). In fact, difficulty tolerating distress (i.e., difficulty persisting in goal-directed activity when experiencing distress) is prominent among youth with conduct problems (Burt, McGue, Iacono, & Krueger, 2006) and adults with APD (Brem, Florimbio, Elmquist, Shorey, & Stuart, 2018; Daughters et al., 2008; Sargeant, Daughters, Curtin, Schuster, & Lejuez, 2011), and is believed to prompt engagement in maladaptive behaviors (e.g., substance use, aggression) to relieve negative affect (Daughters, Gorka, Magidson, MacPherson, & Seitz-Brown, 2013; Daughters et al., 2008; Van Eck, Warren, & Flory, 2017). In the EEfRT, it is possible

that individuals with high levels of antisociality who are experiencing high levels of negative affect have less tolerance for the distress involved in completing effortful tasks, and may thus be less willing to expend effort, regardless of the rewards at stake or their probabilities. Taken together, these findings suggest that negative affect particularly constrains integration of decision variables during effort-based choice among individuals with more severe antisocial behavior.

Another important cost-benefit decision-making factor related to antisociality is delay discounting. Delay-cost sensitivity is found in both youth and adults who engage in antisocial behavior (Petry, 2002; White et al., 2014). Here, we found that participants with steeper delay discounting during inter-temporal choice showed a consistent preference for low effort and low reward options and were less sensitive to information about expected value during effort-based choice. However, this relationship is perhaps not surprising given the overlap in mechanisms held between effort and delay-based discounting (Croxson et al., 2009; Peters & Buchel, 2011; Westbrook, Kester, & Braver, 2013). It is often the case that rewards that require more effort to obtain also are received at greater temporal delays. This is the case in the present task, where the longer duration of the hard task delays the delivery of feedback about reward earnings relative to the easy task. It is possible that the shortsightedness about the future present in individuals who discount delayed rewards more steeply prevents them from appropriately estimating the value of exerting effort to obtain future rewards. In other words, steeper delay discounting may reflect not only a bias towards the present, but also an intolerance or aversion to delays that precede rewards requiring a high amount of effort (Pattij & Vanderschuren, 2008). Although delay discounting was related to effort-based choice,

and often is elevated among antisocial individuals, we did not find evidence that delay discounting moderated the relationship between effort-based computations and antisociality. Follow-up analyses on our EV findings indicated that individuals with more APD symptoms showed reduced sensitivity to multiple decision variables (probability, reward, and their combination, EV), suggestive of integration deficits. By contrast, delay discounting was associated with intact sensitivity to reward magnitude and only diminished sensitivity to probability (see Supplemental Results). This latter finding is consistent with prior studies linking delay-cost sensitivity with abnormal probability-cost sensitivity (Green & Myerson, 2013). More broadly, though, these independent influences on effort-based decision-making reflect how multiple factors can result in the same decision-making aberrations. Overall, these findings highlight the importance of not only considering specific facets of cost-benefit decision-making, but also considering component processes within types of decision costs and individual differences that relate to them (Green & Myerson, 2013). Ultimately, delay- and effort-based preferences may have unique relationships to antisociality and somewhat additively increase vulnerability to diminished cost-benefit decision-making.

The present study is not without limitations. First, with our cross-sectional design, it is unclear whether decreased use of expected value to modulate effort occurs as a *consequence* of more severe antisocial behavior, rather than as a *mechanism* supporting the development of more severe antisociality. Learned industriousness theory (Eisenberger, 1992) suggests that individual differences in exerting effort depend upon previous experiences of reinforcement for effortful behavior. It is possible that individuals with greater levels of antisociality were less likely to integrate information

about reward and probability due to learning from previous experiences being rewarded for low-effort choices (e.g., getting away with not paying bills on time) or insufficiently rewarded for high-effort choices (e.g., not being promoted at a job when they worked hard). Although our effects were robust to various potential confounds (see Supplemental Results), future studies using longitudinal designs are needed to clarify the directionality of the association between effort-based decision-making and antisocial behavior. Second, the goal of the present study was to examine the influence of decision variables on effort-based computations; however, we did not measure effort discounting (Botvinick, Huffstetler, & McGuire, 2009). The EEfRT is designed to look at basic effort-based decision-making, but it does not measure effort discounting, as there was no parametric variation in the amount of effort required over trials. Future studies that sample from a wider distribution of effort options would be useful for testing alterations in effort-based decision-making among individuals who engage in antisocial behavior.

In sum, the present study indicates that individuals with greater levels of antisociality show aberrations in the integration of multiple decision variables to guide action selection during effort-based decision-making. Moreover, these integration deficits appear to be closely linked to states of negative affect and separate from deficits related to delay discounting. Decisions regarding effort define many choices individuals who engage in antisocial behavior are confronted with in the real world. Whether they are deciding to make an effort to get home by curfew, to find a sober driver, or to overcome barriers to reentering the workforce (Visher, Debus-Sherrill, & Yahner, 2011), making adaptive choices requires the integration of information regarding the probabilities of certain outcomes and the potential benefits (e.g., the greater likelihood of getting home

safely, staying out of jail, earning a steady income). Failure to integrate this information can result in choice behavior that yields problematic outcomes (e.g., physical harm, re-arrest) for the individual and other members of society.

**Table: Study 2**

**Table 1**

*Sample characteristics and task statistics*

| | N | Min | Max | M | SD |
|---|---|---|---|---|---|
| Age | 80 | 18 | 62 | 37.59 | 12.88 |
| Sex | | | | | |
| Male | 51 | | | | |
| Female | 29 | | | | |
| Race | | | | | |
| White | 33 | | | | |
| Black | 45 | | | | |
| Asian | 2 | | | | |
| Ethnicity | | | | | |
| Hispanic | 5 | | | | |
| Not Hispanic | 75 | | | | |
| Education | | | | | |
| Junior High/Middle School | 2 | | | | |
| Partial High School | 16 | | | | |
| High School Graduate | 23 | | | | |
| Partial College | 21 | | | | |
| College Education | 11 | | | | |
| Graduate Degree | 7 | | | | |
| Conduct Disorder Diagnosis | | | | | |
| Absent | 54 | | | | |
| Present | 26 | | | | |
| Antisocial Personality Disorder Diagnosis | | | | | |
| Absent | 60 | | | | |
| Present | 20 | | | | |
| Total APD Symptoms | 80 | 0 | 18 | 2.68 | 4.09 |
| Ln (Total APD Symptoms + 1) | 80 | 0 | 2.94 | 0.81 | 0.94 |
| Positive Affect Score | 80 | 15.00 | 50.00 | 33.28 | 8.97 |
| Negative Affect Score | 80 | 10.00 | 31.00 | 13.06 | 4.72 |
| *ln*(Positive Affect Score + 1) | 80 | 2.71 | 3.91 | 3.46 | 0.29 |
| *ln*(Negative Affect Score + 1) | 80 | 2.30 | 3.43 | 2.52 | 0.29 |
| Discount rate *(k)* | 80 | 0.0003 | 0.2500 | 0.0530 | 0.0562 |
| *ln(k)* | 80 | -8.29 | -1.39 | -3.53 | 1.30 |
| EEfRT task – Number of trials | 80 | 44.00 | 50.00 | 49.23 | 1.42 |
| EEfRT task – Proportion of hard task choices | 80 | 0.00 | 1.00 | 0.38 | 0.19 |

## References: Study 2


Barch, D. M., Treadway, M. T., & Schoen, N. (2014). Effort, anhedonia, and function in schizophrenia: reduced effort allocation predicts amotivation and functional impairment. *Journal of Abnormal Psychology, 123*(2), 387.

Black, D. W., Gunter, T., Loveless, P., Allen, J., & Sieleni, B. (2010). Antisocial personality disorder in incarcerated offenders: Psychiatric comorbidity and quality of life. *Annals of Clinical Psychiatry, 22*(2), 113-120.

Bornovalova, M. A., Gratz, K. L., Daughters, S. B., Hunt, E. D., & Lejuez, C. (2012). Initial RCT of a distress tolerance treatment for individuals with substance use disorders. *Drug and alcohol dependence, 122*(1-2), 70-76.

Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 16-27.

Brandon, T. H., Herzog, T. A., Juliano, L. M., Irvin, J. E., Lazev, A. B., & Simmons, V. N. (2003). Pretreatment task persistence predicts smoking cessation outcome. *Journal of Abnormal Psychology, 112*(3), 448.

Brem, M. J., Florimbio, A. R., Elmquist, J., Shorey, R. C., & Stuart, G. L. (2018). Antisocial traits, distress tolerance, and alcohol problems as predictors of intimate partner violence in men arrested for domestic violence. *Psychology of violence, 8*(1), 132.

Buckholtz, J. W. (2015). Social norms, self-control, and the value of antisocial behavior. *Current Opinion in Behavioral Sciences, 3*, 122-129.

Buckholtz, J. W., Karmarkar, U., Ye, S., Brennan, G. M., & Baskin-Sommers, A. (2017). Blunted ambiguity aversion during cost-benefit decisions in antisocial individuals. *Scientific reports, 7*(1), 1-9.

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Benning, S. D., Li, R., . . . Shelby, E. S. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. *Nature neuroscience, 13*(4), 419-421.

Burt, S. A., McGue, M., Iacono, W. G., & Krueger, R. F. (2006). Differential parent-child relationships and adolescent externalizing symptoms: Cross-lagged analyses within a monozygotic twin differences design. *Developmental Psychology, 42*(6), 1289.

Caspi, A., Langley, K., Milne, B., Moffitt, T. E., O'Donovan, M., Owen, M. J., . . . Thapar, A. (2008). A replicated molecular genetic basis for subtyping antisocial behavior in children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 65*(2), 203-210. doi:10.1001/archgenpsychiatry.2007.24

Chong, T.-J., Bonnelle, V., & Husain, M. (2016). Quantifying motivation with effort-based decision-making paradigms in health and disease *Progress in Brain Research* (Vol. 229, pp. 71-100): Elsevier.

Cooper, J. A., Barch, D. M., Reddy, L. F., Horan, W. P., Green, M. F., & Treadway, M. T. (2019). Effortful goal-directed behavior in schizophrenia: Computational subtypes and associations with cognition. *Journal of Abnormal Psychology, 128*(7), 710.

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., & Rushworth, M. F. (2009). Effort-based cost–benefit valuation and the human brain. *Journal of Neuroscience, 29*(14), 4531-4541.

Culbreth, A., Moran, E., & Barch, D. (2018). Effort-cost decision-making in psychosis and depression: could a similar behavioral deficit arise from disparate psychological and neural mechanisms? *Psychological medicine, 48*(6), 889-904.

Daughters, S. B., Gorka, S. M., Magidson, J. F., MacPherson, L., & Seitz-Brown, C. (2013). The role of gender and race in the relation between adolescent distress tolerance and externalizing and internalizing psychopathology. *Journal of Adolescence, 36*(6), 1053-1065.

Daughters, S. B., Sargeant, M. N., Bornovalova, M. A., Gratz, K. L., & Lejuez, C. W. (2008). The relationship between distress tolerance and antisocial personality disorder among male inner-city treatment seeking substance users. *Journal of Personality Disorders, 22*(5), 509-524.

Deater-Deckard, K., Petrill, S. A., & Thompson, L. A. (2007). Anger/frustration, task persistence, and conduct problems in childhood: A behavioral genetic analysis. *Journal of Child Psychology and Psychiatry, 48*(1), 80-87.

Eisenberger, R. (1992). Learned industriousness. *Psychological review, 99*(2), 248.

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). Structured clinical interview for DSM-5—Research version (SCID-5 for DSM-5, research version; SCID-5-RV). *Arlington, VA: American Psychiatric Association*.

Goldstein, R. B., Dawson, D. A., Chou, S. P., Ruan, W. J., Saha, T. D., Pickering, R. P., . . . Grant, B. F. (2008). Antisocial behavioral syndromes and past-year physical

health among adults in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of clinical psychiatry, 69*(3), 368.

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews Neuroscience, 12*(11), 652-669.

Green, L., & Myerson, J. (2013). How many impulsivities? A discounting perspective. *Journal of the Experimental Analysis of Behavior, 99*(1), 3-13.

Gregory, S., Blair, R. J., Simmons, A., Kumari, V., Hodgins, S., & Blackwood, N. (2015). Punishment and psychopathy: a case-control functional MRI investigation of reinforcement learning in violent antisocial personality disordered men. *The Lancet Psychiatry, 2*(2), 153-160.

Hobson, C. W., Scott, S., & Rubia, K. (2011). Investigation of cool and hot executive function in ODD/CD independently of ADHD. *Journal of Child Psychology and Psychiatry, 52*(10), 1035-1043.

Hyde, L. W., Byrd, A. L., Votruba-Drzal, E., Hariri, A. R., & Manuck, S. B. (2014). Amygdala Reactivity and Negative Emotionality: Divergent Correlates of Antisocial Personality and Psychopathy Traits in a Community Sample. *J Abnorm Psychol, 123*(1), 214-224. doi:10.1037/a0035467

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences, 22*(4), 337-349.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *J Exp Psychol Gen, 128*(1), 78-87.

Knapp, M., King, D., Healey, A., & Thomas, C. (2011). Economic outcomes in adulthood and their associations with antisocial conduct, attention deficit and anxiety problems in childhood. *Journal of mental health policy and economics, 14*(3), 137-147.

Kreidler, S. M., Muller, K. E., Grunwald, G. K., Ringham, B. M., Coker-Dukowitz, Z. T., Sakhadeo, U. R., . . . Glueck, D. H. (2013). GLIMMPSE: online power computation for linear models with and without a baseline covariate. *Journal of statistical software, 54*(10).

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology, 114*(4), 537-550. doi:10.1037/0021-843X.114.4.537

Kurniawan, I. T., Guitart-Masip, M., & Dolan, R. J. (2011). Dopamine and effort-based decision making. *Frontiers in Neuroscience, 5*, 81.

Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: a meta-analysis. *Psychological Bulletin, 130*(4), 531-552. doi:10.1037/0033-2909.130.4.531

MATLAB and Statistics Toolbox. (2016b). Natick, Massachusetts.: The MathWorks, Inc.

Matthys, W., Vanderschuren, L. J., & Schutter, D. J. (2013). The neurobiology of oppositional defiant disorder and conduct disorder: altered functioning in three

mental domains. *Development and Psychopathology, 25*(1), 193-207. doi:10.1017/S0954579412000272

Mazas, C. A., Finn, P. R., & Steinmetz, J. E. (2000). Decision-making biases, antisocial personality, and early-onset alcoholism. *Alcoholism: Clinical and Experimental Research, 24*(7), 1036-1040.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *Commons, ML.; Mazur, JE.; Nevin, JA*, 55-73.

Pattij, T., & Vanderschuren, L. J. (2008). The neuropharmacology of impulsive behaviour. *Trends in pharmacological sciences, 29*(4), 192-199.

Peters, J., & Buchel, C. (2011). The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends in Cognitive Sciences, 15*(5), 227-239. doi:10.1016/j.tics.2011.03.002

Petry, N. M. (2002). Discounting of delayed rewards in substance abusers: relationship to antisocial personality disorder. *Psychopharmacology, 162*(4), 425-432.

Phillips, P. E., Walton, M. E., & Jhou, T. C. (2007). Calculating utility: preclinical evidence for cost–benefit analysis by mesolimbic dopamine. *Psychopharmacology, 191*(3), 483-495.

Ponce, G., Jimenez-Arriero, M., Rubio, G., Hoenicka, J., Ampuero, I., Ramos, J., & Palomo, T. (2003). The A1 allele of the DRD2 gene (TaqI A polymorphisms) is associated with antisocial personality in a sample of alcohol-dependent patients. *European Psychiatry, 18*(7), 356-360.

Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., & Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *Journal of Neuroscience, 30*(42), 14080-14090.

Raine, A. (2018). Antisocial personality as a neurodevelopmental disorder. *Annual Review of Clinical Psychology, 14*, 259-289.

Raine, A., Buchsbaum, M., & LaCasse, L. (1997). Brain abnormalities in murderers indicated by positron emission tomography. *Biological Psychiatry, 42*(6), 495-508.

Rosenbloom, M. H., Schmahmann, J. D., & Price, B. H. (2012). The functional neuroanatomy of decision-making. *The Journal of neuropsychiatry and clinical neurosciences, 24*(3), 266-277.

Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. (2006). Separate neural pathways process different decision costs. *Nature neuroscience, 9*(9), 1161-1168.

Salamone, J. D., Correa, M., Farrar, A., & Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology, 191*(3), 461-482.

Salamone, J. D., Correa, M., Yohn, S., Cruz, L. L., San Miguel, N., & Alatorre, L. (2016). The pharmacology of effort-related choice behavior: Dopamine, depression, and individual differences. *Behavioural processes, 127*, 3-17.

Sargeant, M. N., Daughters, S. B., Curtin, J. J., Schuster, R. M., & Lejuez, C. (2011). Unique roles of antisocial personality disorder and psychopathic traits in distress tolerance. *Journal of Abnormal Psychology, 120*(4), 987.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461-464.

Shafiei, N., Gray, M., Viau, V., & Floresco, S. B. (2012). Acute Stress Induces Selective Alterations in Cost/Benefit Decision-Making. *Neuropsychopharmacology, 37*(10), 2194-2209.

Sontag, L. M., Graber, J. A., Brooks-Gunn, J., & Warren, M. P. (2008). Coping with social stress: Implications for psychopathology in young adolescent girls. *Journal of Abnormal Child Psychology, 36*(8), 1159.

StataCorp. (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1): MIT press Cambridge.

Thapar, A., Langley, K., Fowler, T., Rice, F., Turic, D., Whittinger, N., . . . O'Donovan, M. (2005). Catechol O-methyltransferase gene variant and birth weight predict early-onset antisocial behavior in children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 62*(11), 1275-1278.

Treadway, M. T., Bossaller, N. A., Shelton, R. C., & Zald, D. H. (2012a). Effort-based decision-making in major depressive disorder: a translational model of motivational anhedonia. *Journal of Abnormal Psychology, 121*(3), 553.

Treadway, M. T., Buckholtz, J. W., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., . . . Zald, D. H. (2012b). Dopaminergic mechanisms of individual differences in human effort-based decision-making. *Journal of Neuroscience, 32*(18), 6170-6176.

Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E., & Zald, D. H. (2009). Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLoS One, 4*(8), e6598.

Treadway, M. T., & Zald, D. H. (2013). Parsing anhedonia: translational models of reward-processing deficits in psychopathology. *Current Directions in Psychological Science, 22*(3), 244-249.

Van Eck, K., Warren, P., & Flory, K. (2017). A variable-centered and person-centered evaluation of emotion regulation and distress tolerance: links to emotional and behavioral concerns. *Journal of Youth and Adolescence, 46*(1), 136-150.

Visher, C. A., Debus-Sherrill, S. A., & Yahner, J. (2011). Employment after prison: A longitudinal study of former prisoners. *Justice Quarterly, 28*(5), 698-718.

Walton, M. E., Rudebeck, P. H., Bannerman, D. M., & Rushworth, M. F. (2007). Calculating the cost of acting in frontal cortex. *Ann N Y Acad Sci, 1104*, 340.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063.

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLoS One, 8*(7), e68210.

White, S. F., Clanton, R., Brislin, S. J., Meffert, H., Hwang, S., Sinclair, S., & Blair, R. J. R. (2014). Reward: Empirical contribution: Temporal discounting and conduct disorder in adolescents. *Journal of Personality Disorders, 28*(1), 5-18.

Wilkinson, G. S. (1993). *The Wide Range Achievement Test: Manual* (3rd ed.). Wilmington, DE: Wide Range Inc.

Yang, X.-h., Huang, J., Zhu, C.-y., Wang, Y.-f., Cheung, E. F., Chan, R. C., & Xie, G.-r. (2014). Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remitted depression patients. *Psychiatry Research, 220*(3), 874-882.

Yang, Y., & Raine, A. (2009). Prefrontal Structural and Functional Brain Imaging findings in Antisocial, Violent, and Psychopathic Individuals: A Meta-Analysis. *Psychiatry Research, 174*(2), 81-88. doi:10.1016/j.pscychresns.2009.03.012

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised Manual*. Los Angeles, CA: Western Psychological Services.

# Chapter 4: Study 3

## Cognitive remediation as a targeted treatment for cognitive-affective dysfunctions among antisocial individuals

## Abstract

Numerous strategies for curtailing the maladaptive behaviors of antisocial individuals have been proposed, but treatments to date are limited in their effectiveness for behavior change. One limitation of many existing treatments is their failure to consider the cognitive-affective dysfunctions that confer risk for and maintain antisociality. Decades of research demonstrate that antisocial individuals tend to display abnormal inhibition in the presence of rewards, aberrant decision-making amid ambiguity, and deficient working memory under distress. The present study evaluated a novel cognitive remediation package designed to address these cognitive-affective dysfunctions in a sample of individuals enrolled in outpatient substance use treatment. Participants ($N$ = 46) were randomized to either the targeted cognitive remediation training package (CogTrain) or an active control package. Both packages were administered twice per week for 4 weeks in addition to treatment as usual (TAU). Cognitive-affective functioning and real-world behavior were evaluated both pre- and post-training. Results indicated that compared to individuals who completed control training, individuals who completed CogTrain showed advances in working memory, reductions in substance use frequency, and improvements in TAU session attendance after four weeks of training. These improvements correlated with improvements in inhibition efficiency and cognitive persistence amid distress on training tasks and were not impacted by levels of externalizing traits or training dose. These findings suggest that antisocial individuals can benefit from treatments that address the cognitive-affective dysfunctions at the root of their behavior.

Antisocial behavior poses a substantial burden to the healthcare system, the legal system, and society at large, with annual costs in the United States exceeding three trillion dollars when considering expenses related to crime, substance abuse, lost work productivity, and health care (Federal Bureau of Investigation, 2016; Krueger, Markon, Patrick, Benning, & Kramer, 2007; Miller, Cohen, Swedler, Ali, & Hendrie, 2020; National Institute on Drug Abuse, 2017). In response to this enormous burden, many treatments for chronically antisocial individuals have been proposed and tested (Brazil, van Dongen, Maes, Mars, & Baskin-Sommers, 2018; Gibbon, Khalifa, Cheung, Völlm, & McCarthy, 2020; Harris & Rice, 2006; Messina, Farabee, & Rawson, 2003). Despite the vast number and variety of attempted treatments, however, there is limited evidence to suggest that antisocial individuals benefit from treatment (Duggan, Huband, Smailagic, Ferriter, & Adams, 2007; National Collaborating Centre for Mental Health, 2010; Salekin, 2002; Warren et al., 2003). Even Cognitive Behavioral Therapy (CBT), known as the most empirically-supported psychosocial treatment across mental health conditions (Craske, 2010), generally yields null effects as a therapeutic intervention for antisocial individuals (Davidson et al., 2009; Davidson & Tyrer, 1996). Upon further thought, however, this failure of a "gold standard" treatment seems likely when considering the specific and multifaceted cognitive-affective dysfunctions that characterize antisocial individuals and the lack of accommodation for these dysfunctions in standard treatments (Wölwer, Burtscheidt, Redner, Schwarz, & Gaebel, 2001).

Research underscores a significant role for cognitive dysfunction in the presentation of antisocial traits and behaviors (Ogilvie, Stewart, Chan, & Shum, 2011; Raine & Scerbo, 1991). In particular, antisocial individuals show diminished inhibition

(Chamberlain, Derbyshire, Leppink, & Grant, 2016; Dolan & Park, 2002; Rubio, Jiménez, et al., 2007; Swann, Lijffijt, Lane, Steinberg, & Moeller, 2009; Zeier, Baskin-Sommers, Hiatt Racer, & Newman, 2012), a cognitive function responsible for deliberately overriding dominant, automatic, and pre-potent responses to achieve desired goals (Miyake et al., 2000). Antisocial individuals also demonstrate abnormal cost-benefit decision-making, which involves integrating multiple sources and types of information about choice options and their potential outcomes to optimize behavior (Buckholtz, 2015; Buckholtz, Karmarkar, Ye, Brennan, & Baskin-Sommers, 2017; Mazas, Finn, & Steinmetz, 2000; Petry, 2002; White et al., 2014; see also Study 2). Finally, antisocial individuals frequently show reduced working memory capacity (De Brito, Viding, Kumari, Blackwood, & Hodgins, 2013; Gerst, Gunn, & Finn, 2017; Gunn, Gerst, Lake, & Finn, 2018)), reflecting a reduced ability to maintain and update information in an active state in order to achieve specific goals (Purves et al., 2008). These cognitive dysfunctions are especially pronounced in affectively-charged situations (Baskin-Sommers & Newman, 2013): those requiring inhibition in the presence of reward, (Bjork & Pardini, 2015; Buckholtz, 2015; Byrd, Loeber, & Pardini, 2014; Dolan, 2012; Fonseca & Yule, 1995; Ogilvie et al., 2011; Patrick, Durbin, & Moser, 2012; Rowe, 1997), cost-benefit decision-making under ambiguous circumstances (Buckholtz et al., 2017; Mazas et al., 2000; Schönenberg & Jusyte, 2014), and working memory in the context of distress (Brem, Florimbio, Elmquist, Shorey, & Stuart, 2018; Daughters, Sargeant, Bornovalova, Gratz, & Lejuez, 2008; Prehn et al., 2013; Sargeant, Daughters, Curtin, Schuster, & Lejuez, 2011). Importantly, these specific cognitive-affective

dysfunctions can impede antisocial individuals from effectively engaging in therapeutic interventions to change their disruptive behavior patterns (Kim et al., 2018).

In order to benefit from therapy, individuals must attend scheduled therapy sessions and apply insights and skills acquired in treatment to their daily lives (Turner, LaRowe, Horner, Herron, & Malcolm, 2009). Antisocial individuals have difficulty inhibiting strong urges to pursue immediate rewards (e.g., drug cravings, desires to sleep in), and these urges may overshadow any intentions they have to attend scheduled therapy appointments or practice coping skills learned in treatment when faced with temptations. Diminished working memory capability during times of distress also may prevent antisocial individuals from remaining aware of appointment dates and times, their assigned homework, their treatment goals, or the steps for practicing specific coping skills when stressors command their attention. Similarly, a diminished ability to integrate ambiguous cost-benefit information into decisions may make it difficult for antisocial individuals to engage in therapeutic techniques like cognitive restructuring, a core CBT skill that involves carefully considering numerous possible interpretations of ambiguous situations before forming more balanced alternative thoughts and improving emotions and behaviors (Beck, 1985; Clark, 2013). Decision-making abnormalities could lead antisocial individuals to underestimate the unknown costs of continuing problem behaviors (e.g., getting caught and going to jail) and the possible benefits of going to therapy (e.g., passing drug tests for probation or potential jobs), leading them to opt out of treatment altogether. Overall, cognitive-affective dysfunctions associated with antisociality may serve as impediments to attending therapy and are contraindications for many components of standard treatments, which may partly explain why antisocial

individuals are less likely to present to treatment (National Collaborating Centre for Mental Health, 2010), complete treatment (Daughters et al., 2008; Martínez-Raga, Marshall, Keaney, Ball, & Strang, 2002), or benefit from treatment (Compton, Cottle, Jacobs, Ben-Abdallah, & Spitznagel, 2003). Successfully altering the aberrant behavior of antisocial individuals, then, likely depends on successfully targeting the core cognitive-affective dysfunctions that contribute to maladaptive tendencies and poor treatment efficacy for antisocial individuals.

Cognitive remediation is a therapeutic approach that may be well-suited to address the cognitive-affective dysfunctions found among antisocial individuals. Already gaining attention for its efficacy in a wide variety of psychiatric populations (e.g., individuals with schizophrenia, depression; Kim et al., 2018; Millan et al., 2012), cognitive remediation utilizes learning principles derived from basic science to improve upon functional outcomes through improved cognitive functioning (Medalia & Bowie, 2016; Wykes, Huddy, Cellard, McGurk, & Czobor, 2011) in one or more cognitive domains (Kim et al., 2018). Compared to the plethora of research concerning the efficacy of cognitive remediation in other psychiatric populations, studies of cognitive remediation in antisocial individuals have been extremely limited.

In the only existing study of cognitive remediation specifically for antisocial individuals to date, Baskin-Sommers, Curtin, and Newman (2015) developed two computerized cognitive remediation training packages for incarcerated individuals who were awaiting substance abuse treatment. The first was designed to address the cognitive-affective dysfunctions present among individuals high on externalizing, the subtype of antisocial individuals who make up the greatest percentage of incarcerated individuals

and who display chronic antisocial traits, impulsivity, and substance misuse (Estrada, Tillem, Stuppy-Sullivan, & Baskin-Sommers, 2019; Krueger et al., 2009). The second cognitive remediation training package was designed to address the distinct cognitive-affective dysfunctions present among individuals high on psychopathy, a subtype of antisocial individual that also displays impulsive-antisocial behavior, but combined with shallow affect, low prosocial emotions, and grandiosity and is less prevalent than externalizing. For externalizing individuals, the training focused on *affective cognitive control*, and was designed to provide individuals with practice inhibiting behavior (in a neutral or rewarded context) and employing distress tolerance. For psychopathic individuals, the training was focused on *attention to context* and assisted individuals in attending to and integrating contextual cues (neutral or affective) present in the environment. Half of the participants received 6 weeks of cognitive remediation training (3 computerized tasks per once-weekly training) that matched their specific cognitive-affective dysfunctions (i.e., these externalizing individuals received *affective cognitive control* training, psychopathic individuals received *attention to context* training) and half received the training that instead matched the dysfunctions of the other subtype of antisocial individual (i.e., these externalizing individuals received attention to context training, psychopathic individuals received affective cognitive control training). One week after the end of the six-week training period, study participants completed a posttreatment assessment battery that was identical to an assessment battery that was administered pretreatment.

Results from Baskin-Sommers and colleagues (2015) support the hypothesis that targeting the specific dysfunctions of externalizing and psychopathic antisocial subtypes

leads to change in cognitive-affective functioning. For the purposes of the present study, we will focus on externalizing effects. Externalizing individuals who received the dysfunction-matched *affective cognitive control* training improved on trained tasks, demonstrating an enhanced ability to act, rather than over-react, to affective and motivationally salient information. Additionally, externalizing individuals who received dysfunction-matched training demonstrated significant improvement on a separate group of laboratory measures (pre-post tasks) relative to the performance for those who received training that was not matched to their dysfunctions, indicating that matched-training lead to improved cognitive-affective performance more generally.

Though Baskin-Sommers and colleagues (2015) provided the first evidence that it is possible to use cognitive remediation training to target and modify cognitive-affective dysfunctions associated with antisociality, there were aspects of results that call to question the training's potential to ameliorate the cognitive-affective dysfunctions and real-world behavior typical of the majority of antisocial individuals. Specifically, the externalizing-related effect sizes for change on the pre-post tasks were quite small relative to psychopathy-related effect sizes ($\eta_p^2$ of .05 for externalizing individuals who completed dysfunction-matched training, compared to $\eta_p^2$ of .21 for psychopathic individuals who completed dysfunction-matched training). Additionally, much of the externalizing effects for matched relative to non-matched training were attributable to deteriorations in pre-post task performance for those who completed the non-matched (psychopathy-targeted) training, rather than improvements among those who completed the dysfunction-matched training (Baskin-Sommers et al., 2015). Finally, although improvement on cognitive remediation training tasks generalized to improved

performance on separate cognitive-affective measures, changes in meaningful real-world behavioral outcomes such as substance abuse, aggression, or criminal activity were not reported. As such, while cognitive remediation training may be suited to address the cognitive-affective dysfunctions found in antisocial individuals, it remains unclear whether such training produces change that is robust, separable from iatrogenic effects of alternative trainings, and that results in functional changes in real-world behavior.

The goal of the present study was to conduct a pilot randomized controlled trial testing a novel cognitive remediation training package for antisocial individuals. Improving upon the affective-cognitive control training developed by Baskin-Sommers and colleagues (2015), which included training tasks involving cognitive control in the presence of distraction, distress, and incentives, we selected training tasks that targeted a more refined array of cognitive-affective functions related to antisocial (externalizing) individuals: inhibition in the presence of reward, decision-making under ambiguity, and working memory amid distress, based on more recent studies. We chose to test the cognitive remediation package in a sample of individuals in outpatient substance use treatment for two reasons. First, we sought to determine whether cognitive remediation training led to changes in treatment engagement and real-world behavior, and thus needed to study antisocial individuals engaged in treatment outside of a controlled environment (e.g., not in prison or inpatient treatment). Second, while antisocial individuals are unlikely to present in treatment settings more broadly (Glenn, Johnson, & Raine, 2013), outpatient substance use treatment settings are recognized as a context well-suited for reaching antisocial individuals (Gardiner, Tsukagoshi, Nur, & Tyrer, 2010; Thylstrup & Hesse, 2016; Tyrer, Mitchard, Methuen, & Ranger, 2003), and

substance use treatment is linked to decreases in substance use and criminal activity (Ali, Green, Daughters, & Lejuez, 2017). We compared effects of the targeted cognitive remediation training to effects of an active control training that was unlikely to influence antisocial-linked cognitive-affective dysfunctions positively or produce iatrogenic effects. We hypothesized that completing targeted treatment in the new cognitive remediation training package, but not an active control training, would be associated with improvements in cognitive-affective functioning (on training tasks as well as measures of working memory, distress tolerance, and delay discounting) and real-world behavior (substance use frequency, treatment session attendance). We also hypothesized that improvements on performance on the cognitive remediation training tasks themselves would coincide with improvements on broad measures of cognitive functioning and real-world behavior.

## Method

### Participants

Participants were 56 treatment seeking individuals recruited at one of two outpatient substance abuse treatment clinics in New Haven, Connecticut (see Table 1). Both treatment sites provided case management, medication management, and assessment, in addition to individual and group psychotherapy services to community members, many of whom were court-referred to treatment. Individuals at each site were informed by clinicians about ongoing research projects after completing intake appointments and given the option to be contacted by research staff. Clinicians pre-screened interested participants based on background information (i.e., age, diagnoses)

collected at intake. Those who expressed interest in participating in research and appeared eligible based on pre-screenings were invited to meet with study staff to complete screening appointments, consisting of interviews and self-report measures administered at their respective treatment clinics. Individuals were eligible if they: were between the ages of 18 and 50; met the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association, 2013) criteria for a current alcohol, cocaine, cannabis, or opioid use disorder; were not currently physically dependent on opioids or alcohol; had IQ scores estimated to be greater than or equal to 70 on a brief measure of IQ (Zachary, 1986); were estimated to have a fourth grade English reading level or higher; did not meet criteria for schizophrenia, bipolar disorder, or psychosis, not otherwise specified (Structured Clinical Interview for DSM-5 Disorders; First, Williams, Karg, & Spitzer, 2015); were sufficiently stable for 4 weeks of training; had three or fewer head injuries with loss of consciousness for over 30 minutes or lasting effects; had no history of chronic illness or neurological disorders (epilepsy or stroke) that would complicate evaluation of effects of cognitive training; and had no history of certain medical problems (e.g., uncorrectable auditory or visual deficits; head injury with loss of consciousness greater than 30 minutes) that might impact their comprehension of the materials or ability to perform training tasks as they were designed.

As shown in the consort diagram (Figure 1), 56 individuals were provided written informed consent approved by the Yale University Human Investigation Committee and completed screening. Forty-eight (48) of the individuals who completed screening were determined to be eligible for the study and invited to attend future study appointments. Of

these 48, 46[8] individuals presented at their first scheduled training session and were randomly assigned to a training condition via a computerized urn randomization program (Wei & Lachin, 1988) that has been used in multiple previous trials (Ball et al., 2007; Carroll et al., 2008; Carroll et al., 2004; Carroll et al., 2016; Kiluk et al., 2016). The urn randomization program was designed to balance training conditions with respect to gender, age (18-35 vs. 36-50), primary drug (alcohol vs. illicit drugs), last use of primary drug (yes vs. no in past 28 days), and length of time enrolled in outpatient treatment (0-29 days vs. 30+ days). Participants were randomized to complete the Cognitive Remediation Training package plus Treatment as Usual[9] (CogTrain + TAU) or Active Control training plus TAU (Control + TAU). Participants earned $35 gift cards[10] for their completion of the screening interview, pretreatment assessments, and the cognitive-affective battery. Participants were eligible to earn $10 in cash for completing each of 8 training session appointments and additional $35 gift cards for attending the posttreatment session to evaluate change in cognitive-affective functioning using the same cognitive-affective assessment battery administered at pretreatment. To encourage attendance at

[8] An *a priori* power analysis based on effect sizes and observed power in the cognitive remediation study by Baskin-Sommers and colleagues (2015) suggested that a sample size of 100 individuals would allow us to evaluate moderators of treatment response at 80% power. However, we recruited fewer participants than originally planned due to the reduced number of outpatients at study sites who met basic eligibility criteria (i.e., age 18-50, without a diagnosis of schizophrenia, bipolar, or psychosis), the need to share our eligible study pool with another ongoing research project with overlapping eligibility criteria, and having to end data collection early when the research center through which the study was funded closed.

[9] Prior studies examining cognitive remediation in other psychiatric samples indicate that the effectiveness of cognitive remediation is optimized when provided in addition to traditional psychosocial treatment (Kim et al., 2018; Wykes et al., 2011).

[10] All payments above $10 in value were granted in the form of gift cards to local grocery and department stores, due to requests from clinic staff.

computerized training sessions, participants were also eligible to earn an additional $50 gift card bonus for attendance at all 8 training sessions within the 28-day training period.

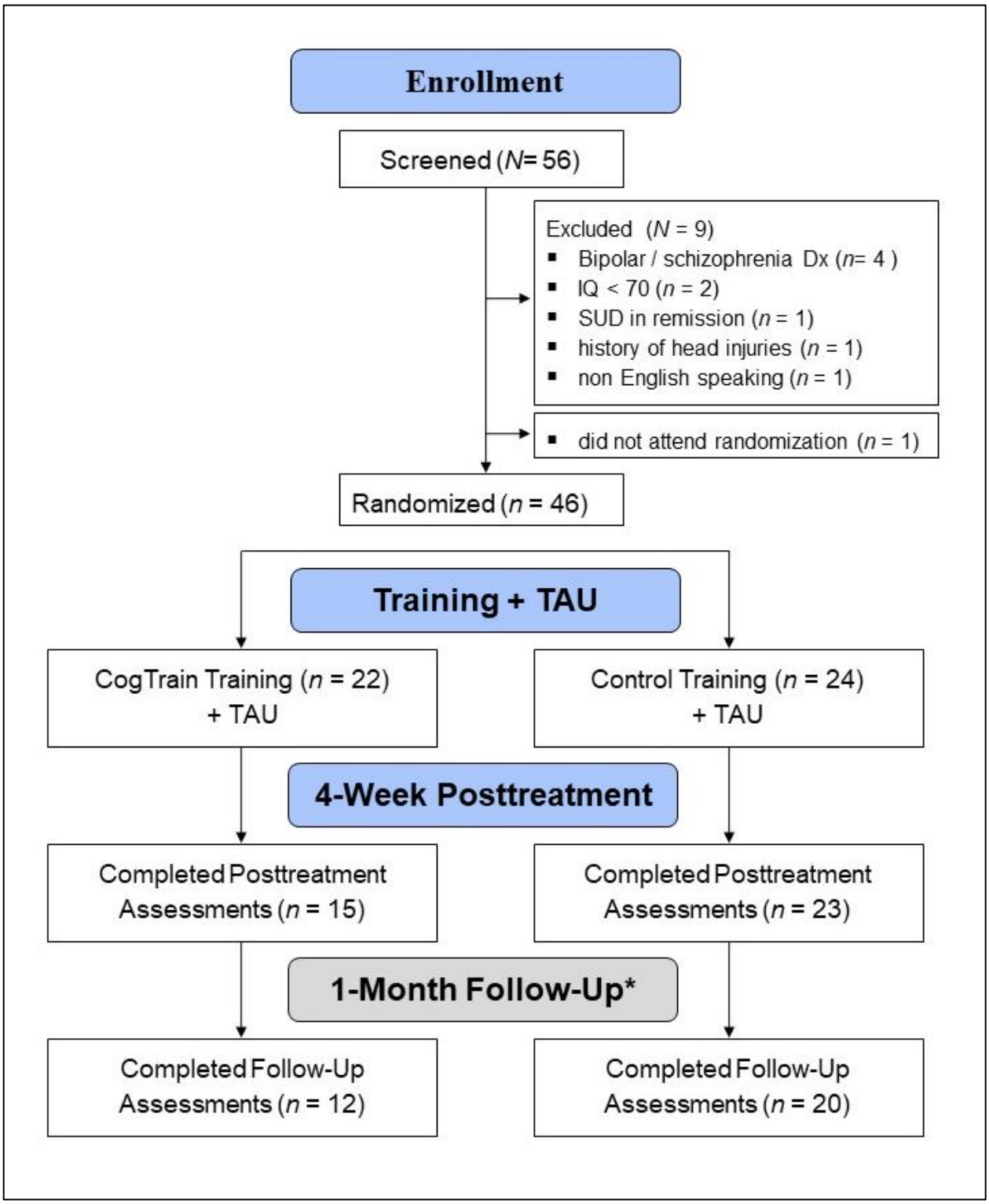


**Figure 5. CONSORT Diagram: Flow of Participants Through the Study Protocol.**
*****_Note:_ Findings for the 1-month follow-up are presented in the Supplementary Material

## Training Conditions

Participants were randomized to one of two training conditions and completed training sessions twice per week for four weeks.[11] All participants concurrently participated in standard treatment, which consisted of a combination of weekly individual counseling sessions and group meetings in addition to case management, medication management, and toxicology screenings.

### *Cognitive Remediation Training (CogTrain)*

Participants randomized into the experimental condition completed the targeted cognitive remediation training package (CogTrain) at each training session. CogTrain provided individuals with experience engaging executive control within affective contexts and consisted of three tasks to address multiple cognitive-affective dysfunctions, rather than any single process. Specifically, tasks tapped inhibition in the presence of reward, working memory in the presence of distress, and cost-benefit decision-making in the presence of uncertainty. We selected three well-validated tasks that reliably tap these processes and that are amenable to repeated administration. Additionally, selected tasks were designed to ensure a broad range of difficulty so that a variety of individuals could find the tasks challenging (Hendershot et al., 2018; Herrera, Chambon, Michel, Paban, & Alescio-Lautier, 2012). Accommodating a wide range of difficulty also decreased the

---

**11** We selected a training period of 4-weeks, since studies examining cognitive remediation training in populations previous studies suggest that over 50% of individuals in substance use treatment drop out within the first month of treatment (Brorson, Arnevik, Rand-Hendriksen, & Duckert, 2013; Stark, 1992), and we wanted a training period that would capture meaningful variability in treatment retention without losing the majority of the sample. We selected 8 total training sessions based on previous studies of cognitive remediation training among individuals in substance use treatment that included between 4 and 12 (Bickel, Yi, Landes, Hill, & Baxter, 2011; Rupp, Kemmler, Kurz, Hinterhuber, & Wolfgang Fleischhacker, 2012) training sessions and produced meaningful cognitive change.

likelihood of ceiling and floor effects, which occur frequently in studies of antisocial individuals (e.g., Dolan & Fullam, 2004; Moody, Franck, & Bickel, 2016) and could have limited our ability to produce and measure cognitive-affective change (Hardy et al., 2015).

**The Stop Signal Task** (Lappin & Eriksen, 1966; Logan & Cowan, 1984; Verbruggen, Logan, & Stevens, 2008; Vince, 1948). The Stop Signal Task was used to train inhibition, in both neutral and rewarded contexts, since the task is a measure of inhibition that can be administered without or with reward, and is influenced by motivational context (Leotti & Wager, 2010). The Stop Signal Task began with two neutral blocks of 64 trials each followed by two rewarded blocks of the same length. Individuals viewed a series of stimuli (squares or circles) and were instructed to make keyboard presses ("Go" responses) to indicate which stimulus they saw (the "f" key for squares, the "j" key for circles) within a certain amount of time (see Figure 2). For a subset (25%) of stimuli, a stop signal (tone) played soon after the onset of the stimulus. On these stop signal trials, individuals were instructed not to press response keys (i.e., they needed to inhibit or cancel their button presses). On reward blocks, individuals earned 5 reward points for each correct response on "Go" trials, and each nonresponse on stop signal trials. A reward of a larger magnitude (5 points vs. 1 point) was included based on results of Study 1, which indicated antisocial individuals have difficulties in self-regulation in the presence of large rewards. For each block type, the first trial for the session began with a stop signal delay of 250ms. A dual staircase tracking procedure was used, such that the task incrementally adjusted the stop signal delay based on individual performance on the stop signal trials. Successful stopping led to a 50ms increase in stop

signal delay on the next trial (making inhibition more difficult), whereas unsuccessful stopping led to a 50ms decrease in stop signal delay on the following trial (making inhibition less difficult). Performance for this task was measured by calculating a stop signal reaction time (SSRT) for each block type (neutral and rewarded) using the established integration method (Logan & Cowan, 1984). According to the integration method, all RTs on Go trials for a given participant's session are arranged in ascending order, and the RT corresponding to the proportion of Stop trials on which inhibition failed is selected. This RT is subtracted from the SSD, providing an estimate of SSRT. This method for calculating SSRT is considered to have less bias compared to other common methods, which are sensitive to positive skew and gradual slowing of RTs, and can increase the likelihood of spurious differences in stopping with and between subjects (Verbruggen, Chambers, & Logan, 2013). In addition to calculating SSRT for neutral and rewarded blocks, we examined scores (i.e., the number of points earned) for rewarded trials at each session, since changes in motivational context can lead to strategic changes in speed-accuracy trade-offs and lead to discrepancies between inhibition efficiency (measured by SSRT) and overall accuracy (measured by total reward points (Herrera et al., 2019; Leotti & Wager, 2010).

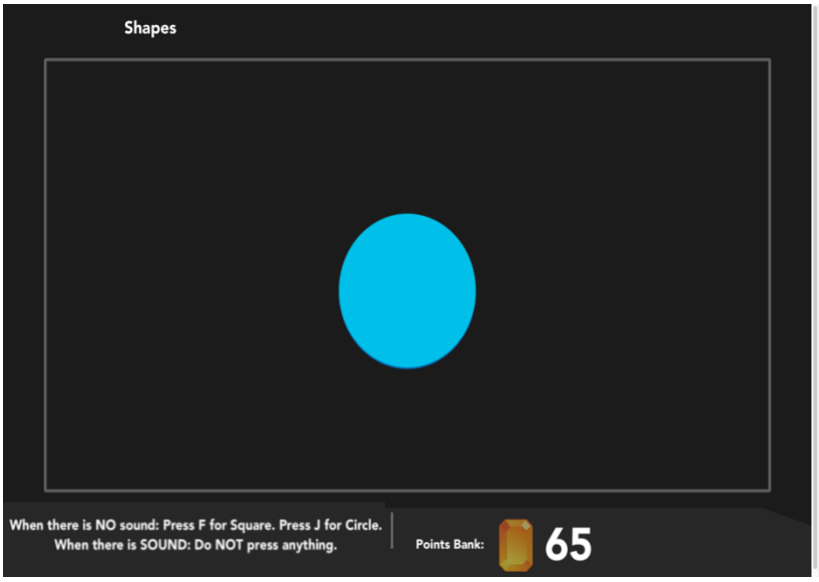


**Figure 6. The Stop Signal Task (rewarded condition).** Participants are instructed to respond indicating whether the presented shape is a circle or a square, unless the shape is presented with a Stop Signal (tone) indicating that participants should withhold a response. In the rewarded condition, participants earn 5 points for each correct trial.

**Decision-Making during Risk and Ambiguity** (adapted from; Konova et al., 2020; Levy, Snell, Nelson, Rustichini, & Glimcher, 2010). This is a decision-making task that aimed to measure and influence ambiguity and risk attitudes. Each trial (of 56 total) presented participants with a distinct virtual "lottery bag" of exactly 100 poker chips (see Figure 3). Within each bag of 100 chips, all chips were colored either red or blue, with the exact number of each color varying from trial to trial. Participants were asked to choose between taking 5 points for certain or playing the lottery (i.e., pulling a chip from the bag) for a chance to win additional points. On 42 “ambiguous” trials, participants received only partial information about the number of red and blue chips in the bag (ranging from having information on 24, 50, or 74 chips in the bag). Participants were informed that the remaining chips, colored in gray, could be either red or blue, and that they would not know the contents of the bag unless they chose to play the lottery. The winning amount for playing the lottery ranged from 6 to 18 points on 39

of the ambiguous trials. In line with previous research, three of the 42 ambiguous trials were “catch” trials, such that the winning amount was equal to that of the certain option (5 points). Catch trials previously have been used in this task to verify that participants understand the task and are adequately attempting to maximize gains (Jia et al., 2020). On the remaining 14 “risk” trials, individuals were given all information about the contents of the lottery bag, and the probability of winning was always 50%, with win amounts also ranging from 5 to 18 points.

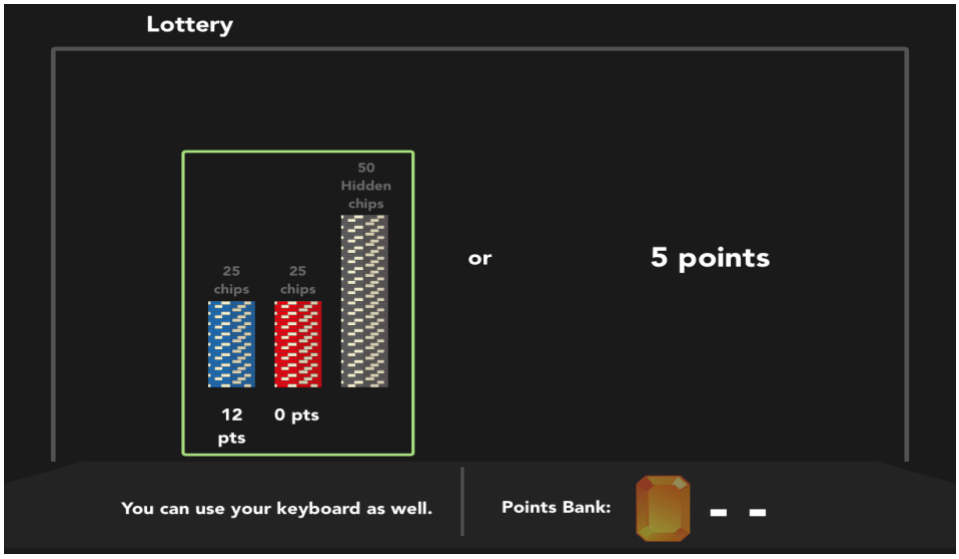


**Figure 3. Ambiguity Decision-Making Task.** Participants are instructed to choose between winning a certain number of points (5) or entering to pull a chip from a lottery bag for a larger number of points. The amount of points available for winning the lottery, the number of winning chips, as well as the amount of information about the contents of the lottery bag, varies from trial to trial. In this example, the lottery can earn a possible 12 points, and 25 of 100 chips in the lottery are known to be winning (red) chips, 25 chips are known to be losing (blue) chips, and 50 chips are unknown (gray) to be winning or losing chips.

Each of the 56 total trials began with a fixation cross, after which participants viewed the available information about the bag's contents and were asked to choose between playing the lottery or getting 5 points for certain. Participants did not have time constraints for responding. Participants were told that they would receive 1 reward point per choice selection, with an optional bonus based on actual choice and outcome during a "randomly" selected trial. At the end of the 56-trial session, the computer selected one ambiguous lottery trial and showed the participant what they chose and what the outcome of the lottery was. To encourage individuals to reconsider unknown information (and potentially behave differently than in future sessions), the computer feedback was rigged for the first three training sessions for a given individual. For each of the first three sessions, if the individual chose the lottery option on over half of the ambiguous lottery trials, the computer selected an ambiguous lottery trial in which the

participant chose the lottery. The computer revealed that the majority of the gray chips for that lottery bag were for the losing color, selected a losing chip, and informed participants that they earned 0 points for their bonus. The computer also noted that if the participant had selected the certain option instead, they would have received a 5-point bonus. By contrast, if the individual chose the certain option on over half of the ambiguous lottery trials for that session, the computer randomly selected an ambiguous lottery trial in which the participant chose the certain option. The computer revealed that the majority of the gray chips for that lottery bag were for the winning color, and informed participants that they would have earned *X* bonus points (always more than the certain amount) had they selected the lottery instead of the certain 5 points. For sessions 4 through 8, one of the ambiguous lottery trials was selected at random, and the colors of the gray chips were revealed to be a random mixture of red and blue chips, with the winning chip being a random selection from the revealed probability distribution for that lottery bag.

Multiple performance measures were calculated to quantify ambiguity and risk preferences on this task. First, we calculated simple measures of choice behavior as the percentage of trials in which the lottery option was selected for both ambiguous and risky trials, since rigged feedback on performance was based on the percentage of ambiguous choices during ambiguity trials. Second, we quantified ambiguity and known-risk attitude using computational modeling methods previously used to examine choice behavior on this task (Konova et al., 2020; Levy et al., 2010). The subjective value (SV) of each option (certain or lottery) on each trial was defined by:

$$\mathrm{SV}_{(\mathrm{option})} = \left[p - \beta\left(\frac{A}{2}\right)\right] v^{\alpha}$$

where *v* was the winning amount (5 points for the certain option, 5-18 points for the lottery option on risk and ambiguity trials), *p* was the objective probability of winning (0 for the certain option, .50 for risk and ambiguity options), *A* was the fraction of *p* that was unknown (0 for the certain and risk options, .26, .50, .76 for ambiguity options), and α and β were subject-specific risk and ambiguity attitude parameters, respectively. A participant with an α of 1 was risk-neutral, less than 1 was risk-averse, and greater than 1 was risk-seeking. A participant with a positive β was ambiguity-averse, behaving as if the gray chips were mostly losing chips, while a participant with a negative β was ambiguity-seeking, behaving as if the gray chips were mostly winning chips.

We estimated α and β by fitting a probabilistic choice function to the trial-by-trial data using maximum-likelihood estimation in MATLAB version R2015a (MathWorks):

$$P_v = \frac{1}{1 + e^{\gamma(SV_L - SV_C))}}$$

where $P_v$ was the probability that the participant chose the lottery, $SV_L$ and $SV_C$ were the SVs of the lottery and certain options, respectively, and γ was another subject-specific parameter representing the slope of the logistic function.

**Paced Auditory Serial Addition Task-Computerized** (PASAT-C; Lejuez, Kahler, & Brown, 2003). The PASAT-C is commonly used as a behavioral measure of distress tolerance and requires individuals to maintain and update information in working memory under distress. For this task, numbers were sequentially flashed on a computer screen, and participants were asked to add each number to the number presented before it before another number was presented (see Figure 4). For example, if the digits '3', '6' and '2' were presented, the participant would need to respond with the correct sum of 3 and 6 ('9') followed by the correct sum of 6 and 2 ('8'). Participants provided answers by

using a computer mouse to click on a number pad displayed on the screen. Participants were told that their score would increase by one point with each correct answer. Incorrect answers or omissions would not affect their total score, but would result in the participant hearing a loud noise blast presented at maximum volume over their headphones. The task consisted of three levels with varying latencies between number presentations. Specifically, the first level of the PASAT provided a 3-s latency between number presentations (i.e., low difficulty) for 2 minutes, the second level provided a 2-s latency for the first 2 minutes and a 1-s latency for the last minute (i.e., medium difficulty), and the third level provided a 1-s latency (i.e., high difficulty) until terminated, up to 7 minutes. Performance on this task is traditionally indexed as latency in seconds to task termination. However, the present sample reached the maximum latency to termination on 75% of completed sessions, which is consistent with modal responses observed in prior studies (e.g., McHugh et al., 2011). With the majority of participants being exposed to all 278 level three trials, we opted to measure the extent to which individuals actively engaged in goal-directed behavior (by providing a response on the number pad) during the third level, rather than measure the duration of time that they passively viewed trials in level three. Thus, PASAT-C performance was measured as a percentage of presented level three trials on which responses were attempted (i.e., in which the participant made a selection on the keypad). This measure was normally distributed across participants.

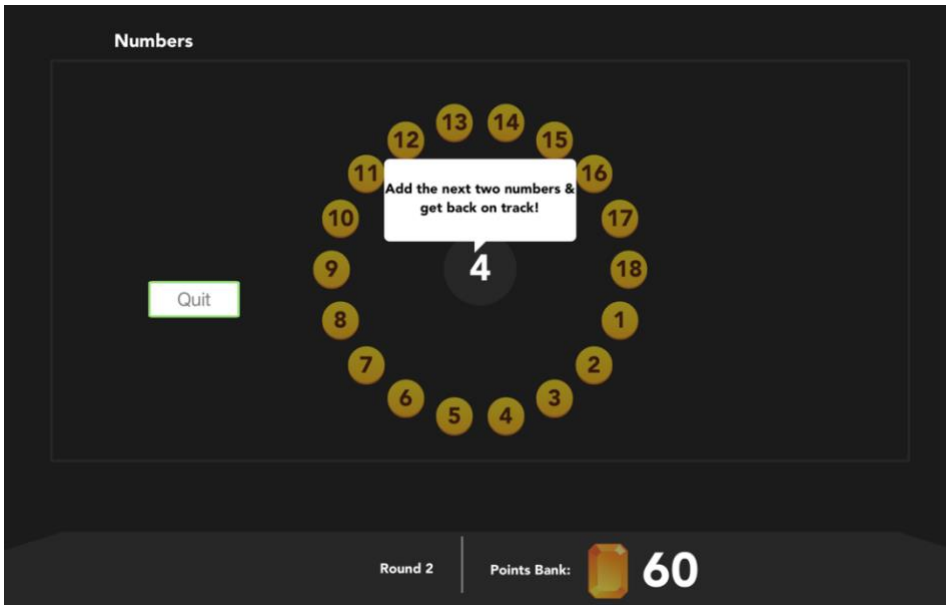


**Figure 4. The Paced Auditory Serial Addition Task - Computerized (PASAT-C).** A screen capture from the third round of the PASAT-C, which features a Quit button that allows participants to discontinue the task. Throughout the task, participants are instructed to add the letter presented in the center of the screen to the letter presented previously and enter the sum using the numbered buttons on the keypad before the next letter is presented.

### *Active Control Training*

**Text Twist.** Participants randomized to this condition completed a computerized word game (Text Twist) at each training visit. Participants were presented with several sets of seven letters and were instructed to make as many words as possible with the letter combinations in the time allotted (see Figure 5). Participants received reward points for every dictionary word identified. An active computer-based control was selected as it provides a rigorous comparison to intervention (Bickel et al., 2011) and addresses concerns that any game-style activities would produce improvements by controlling for computer time and game experiences, as well as non-specific elements such as support, attention, study contact, activation, and motivation. It is possible an active and engaging control condition could produce modest neurocognitive change in some domains, but available data indicate the control condition is not likely to influence the specific

cognitive-affective processes identified in antisocial individuals (Gyurak, Gross, Chan, & Etkin, 2013).

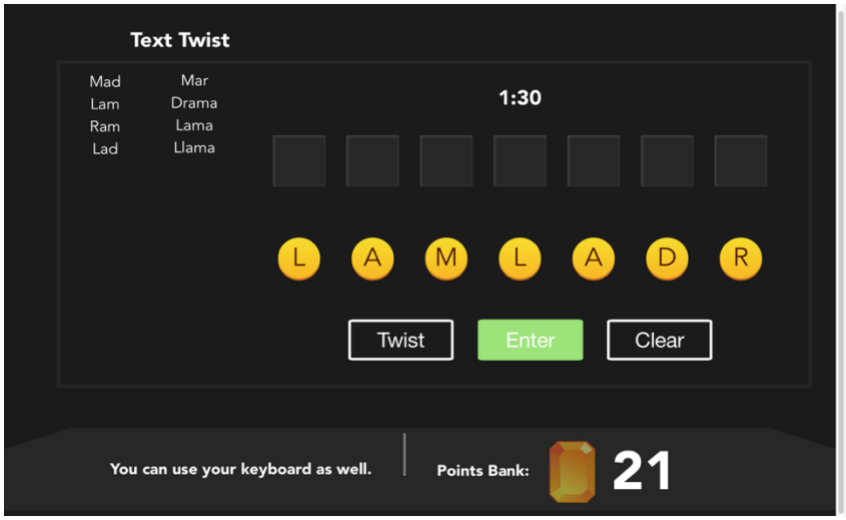


**Figure 5. Active Control Training (Text Twist).** For each trial, participants view 7 letters and are instructed to form as many two to seven letter dictionary words as possible within 2 minutes.

### *Training Session Procedures*

For the first two training sessions, participants in either training condition viewed full computerized instructions for training tasks, with a research assistant reading instructions out loud and being available to answer questions. For sessions 3 through 8, participants were permitted to skip the full instructions and view abbreviated instructions before beginning the tasks. For all 8 training sessions in each condition, participants wore noise-cancelling headphones to reduce distractions and ensure they heard tones for tasks with necessary audio components. At the end of each training task, participants learned their most recent score and viewed the current session score on a graph with all their previous scores for the task. Research assistants commented on whether their score had increased, decreased, or stayed the same since the last session, and noted that they were improving on (in the case of a score that increased) or practicing (in the case of scores that decreased or stayed the same, or scores for session 1) the particular skill targeted by

the training task. An explanation of the skill relative to real-world behavior was provided to illustrate how cognitive training related to participants outside of the training task (e.g., for the PASAT: "*Practicing [the PASAT] may help you keep going when things get stressful, whether in the game itself or in real life. For example, your goal to remain abstinent may become harder when you start to experience stress, withdrawal, or difficult emotions. Practicing this skill during the Numbers game may help you continue practicing abstinence under stress, or may help you get back on track after making mistakes,*" for the active control task, "*In this game, you are working on paying attention for a long period of time and improving your verbal skills. Both of these are skills we use every day when we are reading directions or talking to people around us*").

## Measures

### *Pretreatment Assessments*

**The Externalizing Spectrum Inventory** (The ESI-Brief; Hall, Bernat, & Patrick, 2007). Antisociality is prevalent among individuals in substance use treatment (Thylstrup & Hesse, 2016). Previous studies examining substance use treatment response indicate poorer outcomes among antisocial individuals (Compton et al., 2003). To examine whether training task performance or training response differed as a function of level of trait antisociality, we included the ESI-Brief, a 100-item self-report questionnaire developed to assess a broad range of behaviors and personality associated with externalizing psychopathology (i.e., antisocial personality disorder, substance use disorders, low constraint).

**The State-Trait Anxiety Inventory** (STAI; Kendall, Finch, Auerbach, Hooke, & Mikulka, 1976). Compared to other individuals with substance use disorders, antisocial

individuals have higher anxiety severity (Ducci et al., 2007; Hatzitaskos, Soldatos, Kokkevi, & Stefanis, 1999), which is associated with greater functional impairment and negatively impacts both engagement in and recovery from treatment (Buckner & Carroll, 2010; Thevos, Thomas, & Randall, 1999; Thomas, Thevos, & Randall, 1999). To examine the role of anxiety and its role in treatment engagement and response among antisocial individuals in substance use treatment, we administered the STAI, a widely used measure of consistent (trait) and transient (state) stress and anxiety.

**The Perceived Stress Scale** (PSS; Cohen, Kamarck, & Mermelstein, 1983). Psychological stress occurs when an individual perceives that environmental demands tax or exceed their adaptive capacity. The PSS provides a measure of how much an individual appraises situations in their life as stressful and was administered to examine the relationships among perceived stress, treatment attendance, and response. PSS scores are believed to index a higher-order latent internalizing disposition, or a tendency to experience negative symptoms that are focused inward (Krueger, McGue, & Iacono, 2001). Higher PSS scores are associated with elevated substance use and relapse following substance use treatment (Berg et al., 2010; Lijffijt, Hu, & Swann, 2014; Roos, Kiluk, McHugh, & Carroll, 2020; Tomlinson, Tegge, Athamneh, & Bickel, 2020).

**The Circumstances, Motivation, and Readiness Scale** (CMR; De Leon, Melnick, Thomas, Kressel, & Wexler, 2000). Low motivation and low readiness to change are associated with early dropout from substance abuse treatment, and are core characteristics of antisocial individuals (Ball, Carroll, Canning-Ball, & Rounsaville, 2006). To examine the extent to which individuals' motivation and willingness to enter treatment influenced their responses to cognitive training, we administered the CMR for

Substance Abuse Treatment, an 18-item self-report measure that assesses the circumstances surrounding treatment attendance (e.g., pressure to enter and leave treatment), individuals' motivations for seeking treatment (e.g., their perceived need to change), and individuals' readiness to engage in treatment (e.g., the extent to which they think treatment is necessary to change). Overall, CMR scores serve as an indication of participants' potential motivation and willingness to enter treatment.

### *Cognitive-Affective Assessments*

To examine whether targeted cognitive remediation training resulted in changes to cognitive-affective functioning outside of performance on training tasks, a cognitive-affective battery was administered at pretreatment and posttreatment. This battery consisted of three assessments that have been used previously to tap cognitive-affective processes associated with antisociality.

**Digit Span Backwards** (Wechsler, 1945). Poor working memory performance is common among antisocial individuals (De Brito et al., 2013; Dolan & Park, 2002). The Digit Span Backwards test is a 7-item measure from the Wechsler Digit Span Test within the Wechsler Adult Intelligence Scale. Individuals were presented with progressively longer strings of numbers and were asked to repeat them in reverse order. Each item consisted of two strings of digits, starting with two three-digit strings. To progress from one item to the next, individuals needed to respond correctly on at least one of two strings in an item. Individuals received one point for each correct digit string. The Digit Span Backwards test measured each participant's ability to examine, manipulate, and relay information being held in short-term memory.

**The Breath Holding task** (Daughters, Lejuez, Bornovalova, et al., 2005; Hajek, Belcher, & Stapleton, 1987; Sutterlin et al., 2013; Zvolensky, Feldner, Eifert, & Brown, 2001). Antisocial individuals show reduced persistence in goal-directed behavior in the face of emotional distress on both cognitive and behavioral measures of distress tolerance (Brem et al., 2018; Daughters et al., 2008; Sargeant et al., 2011), and this tendency is linked to a higher likelihood of treatment dropout (Daughters, Lejuez, Bornovalova, et al., 2005; Daughters, Lejuez, Kahler, Strong, & Brown, 2005). The breath holding task was administered to measure behavioral distress tolerance. On this task, participants were instructed to hold their breath for as long as they could and to notify research assistants (by hand raise) when they first began to feel uncomfortable. Research assistants recorded each participant's start time, the time at which they reported that they were feeling uncomfortable, and the time at which participant exhaled. Persistence in the face of discomfort (i.e., distress) was calculated as the difference between the ending time and the time at which participants first reported feeling uncomfortable

**The Five-Trial Adjusting-Delay Discounting Task** (Koffarnus & Bickel, 2014). Antisocial individuals are more likely to choose small immediate rewards over larger rewards available at a delay, reflecting high levels of delay discounting (Petry, 2002). The Five-Trial Adjusting-Delay Discounting Task is a 5-item behavioral assessment used to measure delay discounting. Five questions asked whether the individual would prefer to receive $50 today or $100 at a delay, with the delay duration adjusting based on participant choices. The final delay (i.e., $ED_{50}$) indicated the delay where the current value of the delayed reward ($100) was equal to half of its value ($50) and was also equal to 1/*k*, with *k* being the discount rate for the delayed reward calculated according to a

hyperbolic discounting function (Mazur, 1987), $V_d = r / (1 + kD)$, where $V_d$ was the subjective value of a delayed reward of magnitude *r* available at delay *D*. Distributions of *k* estimations were natural log-transformed due to positive skew. Higher ln*k* values reflected higher levels of delay discounting.

***Real-World Behavior***

To examine whether targeted cognitive remediation training resulted in changes in real-world behavior, interviewers assessed substance use at each study visit and assessed treatment session attendance at pretreatment and posttreatment.

**Substance Use Calendar.** Substance use for each participant's primary drug of choice (cocaine, alcohol, opioids, or marijuana) was documented at each contact via the Substance Use Calendar. Similar to the Time Line Follow-Back (Robinson, Sobell, Sobell, & Leo, 2012), the Substance Use Calendar is an interview assessment administered by a research assistant. Substance use was documented on a daily basis throughout the 28-day training period as well as for the 28-days before random assignment. Substance use calendar data was corroborated using urine and breath screenings at each contact. The urine samples showed excellent correspondence with participants' self-reports of recent substance use with only 6 of 403 urine samples (1.4%), indicating primary drug substance use when the participant denied recent use. Research assistants confronted participants about discrepancies between self-reported substance use and urine results. Participants corrected their responses to substance use calendar questions on four out of the six occasions in which these discrepancies were identified. None of the 146 breath samples indicated alcohol use when a participant with primary alcohol use denied recent alcohol use. Total days of primary substance use (out of 28) for

both the 28-day training period as well as the 28-day pre-training period were calculated and cube-root transformed due to positive skew to provide a measure of primary drug use frequency before and during the training period.

**Treatment as Usual Session Attendance.** Antisocial individuals are at high risk of dropout in outpatient substance use treatment (Thylstrup & Hesse, 2016). One goal of targeted treatment was to enhance engagement in TAU services. TAU at both treatment sites consisted of individual treatment, group treatment, or a combination, based on intake assessment of clinical need and existing mental health care treatment, with some individuals being dually-treated by outside therapists. To examine the impact of cognitive remediation training on treatment engagement, the number of attended individual and group treatment sessions at both the enrolled outpatient treatment clinic and external treatment sites was summed to provide a measure of treatment session attendance at pretreatment and posttreatment for each participant.

## Results

### Sample Description

Table 1 presents baseline demographic characteristics, psychiatric diagnoses, and other assessment measures for the 46 participants (Mean Age = 33.6 years, SD = 8.36) who attended the first training session. Of these, 48% were women, 63% identified themselves as White, 24% identified themselves as Black, 13% did not report their race, and 26% identified their ethnicity as Hispanic. Most were single or divorced (80%), 46% were unemployed, and 93% had completed high school or obtained a GED.

Most participants reported alcohol use (33%) or opioid use (33%) as their primary substance use problem, followed by marijuana (20%), and cocaine (15%). Participants reported 13 years of use for their primary drug on average (SD = 8.0 years), and the majority of participants (85%) experienced severe impairment in functioning based on SUD symptom counts for their primary drug. Consistent with prior studies of antisocial individuals (Brennan, Stuppy-Sullivan, Brazil, & Baskin-Sommers, 2017), problematic use of multiple substances was common, with 85% of participants meeting criteria for two or more substance use disorders in their lifetimes. While only one individual in each training condition met criteria for antisocial personality disorder, levels of trait externalizing (measured with the ESI-Brief) among individuals in our sample were comparable to those found in previous studies examining incarcerated antisocial individuals (e.g., Brennan et al., 2017).

Analysis of variance tests indicated significant differences in baseline distress tolerance scores (as measured by the Breath Holding task) and TAU attendance by treatment condition. Specifically, individuals in the CogTrain condition had lower distress tolerance, lower baseline TAU attendance, attended fewer training sessions, and were less likely to attend the posttreatment session compared to individuals in the control condition. No other analyses examining differences in demographic and baseline assessment measures indicated significant differences between training conditions (see Table 1).

**Study Session Attendance**

Participants completed an average of 6.2 of 8 possible training sessions (SD = 2.3; see Table 2). Of the 46 participants randomized into a training condition, 83% attended

the posttreatment session. Analysis of variance and chi-square tests indicated significant differences in number of training sessions attended and likelihood of returning for the posttreatment assessment battery by training condition, with individuals in the CogTrain condition showing reduced attendance at training and posttreatment sessions compared to individuals in the control condition.

**Training Effects**

For the CogTrain condition, the 22 randomized participants provided data for a total of 120 sessions of the Stop Signal and decision-making tasks, and for 114 sessions of the PASAT-C (one participant refused to complete the PASAT-C due to discomfort hearing the noise bursts, and provided 6 sessions each for the other two tasks in the training package). For the control condition, the 24 randomized participants provided data for 165 sessions of Text Twist. To ensure individuals understood training tasks and met the necessary assumptions for calculating various performance metrics, we excluded session data on a session-by-session basis based on recommendations from prior studies using the training tasks (see Supplementary Material for details). We examined change in performance over time for each training task to see if performance on cognitive training tasks improved with repeated practice. We began by extracting key variables for each task (SSRT for both trial types and total score for reward blocks for each session of the Stop Signal Task; the percentage of risky and ambiguous lottery choices, as well as estimates of $\alpha$ and $\beta$, for the decision-making task; the percentage of third level trials on which participants provided a response during the PASAT-C; and the total number of completed dictionary words for Text Twist; see Table 3 for descriptive statistics and zero-order correlations of training task variables at the first training session for each

task). We included data for each performance variable in a separate mixed effects linear regression using IBM SPSS Statistics (version 26). For each model, the key variable in question was considered as a continuous outcome variable, and session number was included as a continuous fixed-effect predictor. Participant was treated as a random effect.

***CogTrain Training Effects***

Consistent with prior research indicating improvement over time on the Stop Signal Task (Berkman, Kahn, & Merchant, 2014; Manuel, Bernasconi, & Spierer, 2013), there was a significant main effect of session number on SSRT for neutral trials (B = -7.74, SE = 2.86, 95% CI = -13.43, -2.05, , $p$ =.008; see Figure 6A), such that participants exhibited lower SSRT's (i.e., more efficient inhibition) for neutral trials with repeated sessions. The effect of session number on SSRT for reward trials indicated a trend (B = -2.84, SE = 1.62, 95% CI = -6.05, 0.38, $p$ =.084; see Figure 6B) in the same direction. Moreover, there was a significant session number effect on Stop Signal Task score for reward blocks (B = 2.93, SE = 1.25, 95% CI = 0.45, 5.41, $p$ = .021; see Figure 6C), such that accuracy on reward trials improved over repeated sessions.

On the Decision-Making during Risk and Ambiguity task, there was a significant main effect for session number (B = -0.011, SE = 0.006, 95% CI = -0.023, -0.001, $p$ = 0.045; Figure 6D) on percentage of ambiguous choices, such that individuals made fewer choices to play the lottery on ambiguity trials over time. The main effect for session number on percentage of risky choices was not significant (B = -0.004, SE = 0.007, 95% CI = -0.018, -0.009, $p$ = 0.514), indicating choices to play the lottery on risk-only trials did not systematically change across sessions. For parameters derived from

computational models quantifying risk and ambiguity attitudes, the effects for session number were not significant (α: B = 0.031, SE = 0.021, 95% CI = -0.010, 0.072, $p$ = 0.136; β: B = 0.015, SE = 0.037, 95% CI = -0.058, 0.088, $p$ = 0.687).

On the PASAT-C measure of distress tolerance, there was a significant effect for session number (B = 0.043, SE = 0.006, 95% CI = 0.031, 0.054, $p < 0.001$; Figure 6E), with individuals making active attempts to respond during the most difficult level on a greater percentage of trials with repeated sessions.

***Control Training Effect***

There was a significant main effect for session number (B = 13.06, SE = 0.93, 95% CI = 11.22, 14.90, $p < 0.001$; Figure 6F) on number of completed dictionary words, such that performance for the control training task improved across sessions.

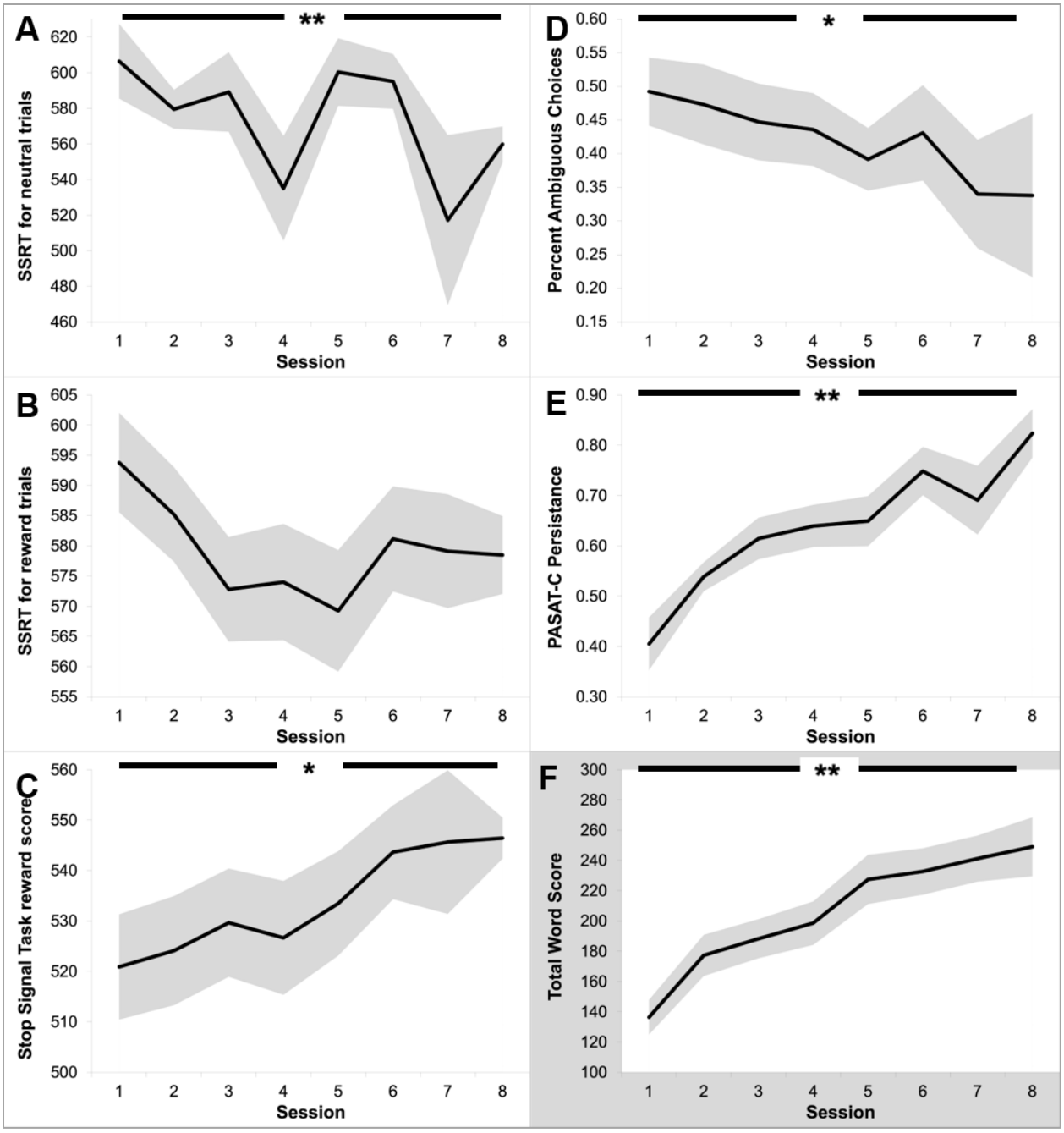


**Figure 6. Training effects for each training task.** On the cognitive remediation training tasks, and the active control training task, there were significant changes in performance on key variables, suggesting performance improved over the course of training. There was a significant effect of session number on SSRT for neutral trials (A) such that inhibition improved across sessions. The effect of session number on SSRT reward trials (B) indicated a trend ($p = .084$) for improvement in inhibition for rewarded trials. There was a significant effect of session number on Stop Signal Task scores for rewarded trials (C), with higher scores obtained on rewarded trials across sessions. On the decision-making task, there was a significant effect of session number on percentage of ambiguous lottery choices (D), with fewer ambiguous choice selections across sessions. On the PASAT-C, there was a significant effect of session number on persistence in the final round (E). Finally, for participants in the active control condition, there was a

significant effect of session number on words completed in the Text Twist task.
**p < 0.01, *p > .0.05

### *Comparing Change Between Conditions*

To compare improvement across trained tasks by training condition, we calculated change scores for each key variable that showed significant improvement across sessions for each participant (SSRT for neutral trials and Total Score for reward trials on the Stop Signal Task, percentage of ambiguous choices on the decision-making task, persistence scores on the PASAT-C, and Total Word scores on Text Twist). Each of the values for a given training task key variable was regressed on session number to create a beta value to capture training change for a given participant's performance on that variable. Performance metrics were transformed such that higher beta values represented greater improvement over training (i.e., lower SSRTs and higher scores on rewarded trials of the Stop Signal Task), decreased ambiguous choices during decision-making, greater persistence during the PASAT-C, more completed words on Text Twist; see Table 4).

For the three CogTrain tasks, we conducted a principal component analysis (PCA) to determine whether improvement across the three tasks reflected improvement on a single factor. We entered Beta values for the four key variables on which performance improved across sessions and selected orthogonal rotation (varimax). The result was a two-factor solution, with Bartlett's test of sphericity indicating that correlations between items were not sufficiently large for PCA, $\chi^2$ (6) = 1.39, $p$ = .967. We, therefore, opted to run separate analyses comparing effects for each of the three CogTrain tasks to effects for the control (Text Twist) task.

A one-way Analysis of Variance (ANOVA) was conducted, with training condition as a between-subjects factor. There was a significant effect of training condition on Beta value when comparing improvement on SSRT for neutral trials to improvement on Text Twist, $F(1, 40) = 32.66$, $p < .001$, $\eta^2 = .45$, with less improvement on inhibition on the Stop Signal Task neutral trials compared to improvement on word scores for Text twist. The effect of training condition also was significant when comparing Beta values for Stop Signal Task reward scores to Text Twist scores, $F(1, 40) = 18.82$, $p < .001$, $\eta^2 = .32$, with less improvement on Stop Signal Task reward trials compared to improvement on word scores for Text Twist. The effect of training condition was not significant when comparing Beta values between the decision-making task and Text Twist, $F(1, 41) = 2.18$, $p = .147$, $\eta^2 = .05$, with no difference in improvement on ambiguous choices compared to improvement on control task performance. Finally, there was a trend for a group effect when comparing Beta values between the PASAT-C distress tolerance scores and Text Twist, $F(1, 40) = 3.93$, $p = .054$, $\eta^2 = .09$, suggesting more improvement on persistence on the PASAT-C compared to improvement on word scores for Text Twist. Overall, there was significant training change for CogTrain and control tasks, with relative change between conditions varying with the CogTrain task variable in question.

## Generalizability of Effects

### *Covariate Analyses*

We next sought to examine whether completing CogTrain versus control training led to generalizable change in cognitive-affective functioning and real-world behavior. We first conducted a series of correlational analyses examining the relation between a

variety of variables (i.e., demographic, baseline) and cognitive-affective and real-world behavior variables to determine which variable(s) to include as covariates in subsequent analyses. Higher IQ was associated with better working memory on the Digit Span Backwards test, $r(37) = .431$, $p = .006$, higher distress tolerance scores during the breath holding task, $r(44) = .433$, $p = .003$, and lower discount rates on the minute discounting task, $r(44) = -.382$, $p = .009$. Additionally, older age was associated with decreased primary drug substance use on the substance use calendar, $r(44) = -.310$, $p = .036$. Therefore, we included full scale IQ in analyses of working memory, distress tolerance, and delay discounting, and Age in analyses of primary drug substance use frequency.

***Pre-post change by training condition***

Next, we examined change in performance on the pre-post cognitive-affective and real-world behavior measures over time as a function of training condition (CogTrain versus control training).[12] For each analysis, we conducted a repeated measures ANOVA with timepoint (pretreatment vs. posttreatment) included as a within-subjects factor and training condition (cognitive remediation vs. active control) included as a between-subjects factor. For analyses of working memory, distress tolerance, and delay discounting performance, full scale IQ (*z*-scored) was included as a continuous covariate, and for the analysis of primary substance use, age (*z*-scored) was included as a continuous covariate.

**Cognitive-Affective Performance**

***Working Memory (Digit Backwards)***. There was a significant main effect of timepoint on working memory, $F(1, 29) = 5.62$, $p = .025$, $\eta^2 = .16$, with participants

[12] We present complete-case analyses in the main text. For analyses using multiple imputation for missing posttreatment values, see supplementary materials.

showing better working memory performance on Digit Backwards at posttreatment compared to pretreatment. The main effect of training condition was not significant ($p$ = .494). However, as predicted, a significant interaction between timepoint and training condition occurred, $F(1, 29) = 21.1$, $p = .007$, $\eta^2 = .22$ (see Figure 7A), such that participants assigned to complete CogTrain training demonstrated improved working memory ability from pretreatment to posttreatment ($p = .002$), while participants assigned to complete Control training did not show change in their working memory ability ($p$ = .654).

***Distress Tolerance.*** On the breath holding task, the main effect of timepoint on distress tolerance was not significant ($p = .883$). The main effect of training condition indicated a trend, $F(1, 35) = 4.05$, $p = .052$, $\eta^2 = .10$, with individuals assigned to the CogTrain training condition displaying lower levels of distress tolerance (consistent with the one-way ANOVA showing a difference in distress tolerance between conditions). The interaction between timepoint and training condition also indicated a trend in the predicted direction, $F(1, 35) = 3.31$, $p = .078$, $\eta^2 = .09$, suggesting that distress tolerance tended to improve from pre- to posttreatment among participants in the CogTrain condition but tended to decrease from pretreatment to posttreatment among participants in the active control training condition.

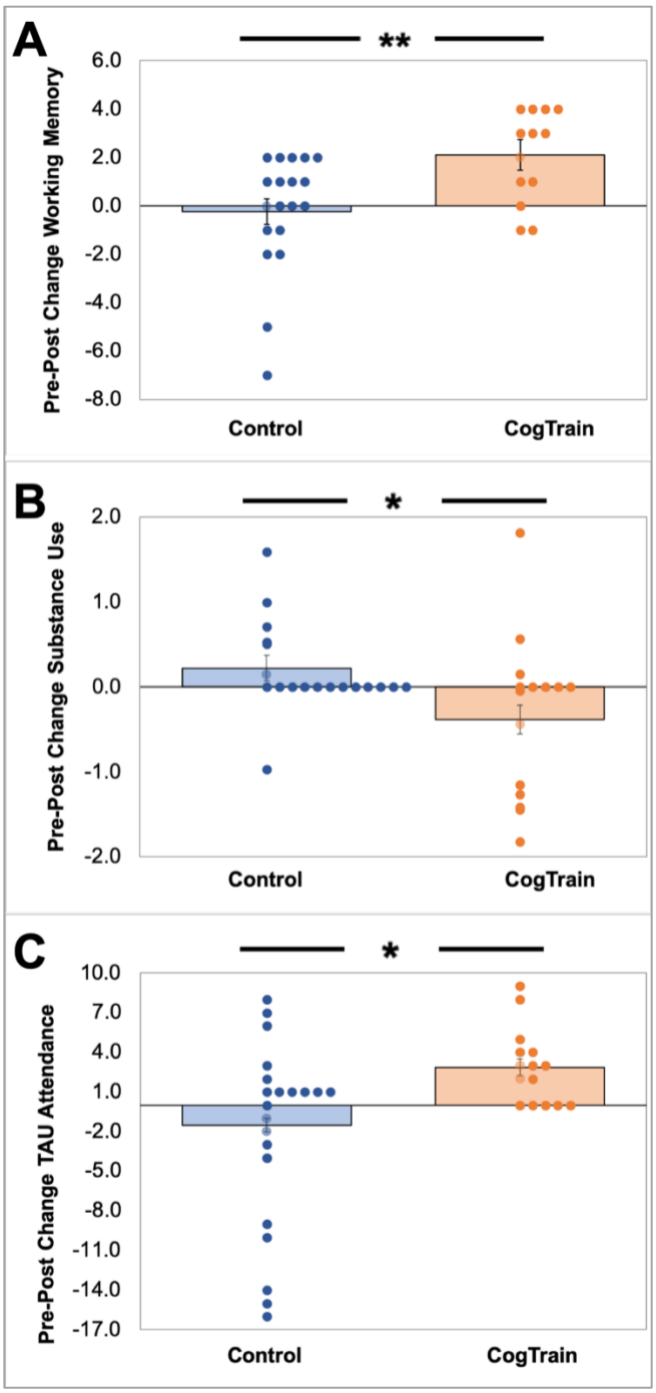


**Figure 7.** Generalizability of effects on measures of cognitive-affective functioning and real-world behavior. There were significant timepoint by training condition effects for working memory, primary drug substance use frequency, and TAU attendance. Specifically, participants who completed targeted cognitive remediation training (CogTrain) showed (A) improvement in working memory ability on the digit backwards task (B) reductions in primary drug substance use frequency, and (C) increased TAU session attendance from pretreatment to posttreatment, whereas participants who completed active control training showed no change in these variables over the same period.
**p < 0.01, *p > .0.05

***Delay Discounting.*** For delay discounting rates, neither the main effect for timepoint ($p$ = .104) nor the main effect of training condition ($p$ = .412) was significant. Similarly, the interaction between timepoint and training condition indicated no significant change in delay discounting between training conditions ($p$ = .395).

**Real-World Behavior**

***Substance Use.*** For primary drug substance use frequency, neither the main effect of timepoint ($p$ = .778) nor the main effect of training condition ($p$ = .270) was significant. There was, however, a significant interaction between timepoint and training condition, $F(1, 31) = 7.17$, $p$ = .012, $\eta^2$ = .19 in the predicted direction (see Figure 7B). Among individuals in CogTrain training, there was a significant decrease in primary drug substance use frequency ($p$ = .030) after 4 weeks of training. By contrast, individuals in control training did not show significant change in primary drug substance use from pretreatment to posttreatment ($p$ = .146).

***TAU Session Attendance***. The main effect of timepoint was not significant ($p$ = .497) for treatment session attendance. The main effect of training condition indicated a nonsignificant trend, $F(1, 36) = 3.52$, $p$ = .069, $\eta^2$ = .09, such that individuals assigned to CogTrain tended to attend fewer treatment sessions (consistent with the one-way ANOVA showing a between-group difference in TAU session attendance between groups at baseline). However, there was a significant interaction between timepoint and training condition, $F(1, 36) = 5.02$, $p$ = .031, $\eta^2$ = .12, indicating that as predicted, individuals who completed CogTrain attended more TAU sessions from pretreatment to posttreatment, whereas individuals who completed Control training did not show change in TAU session attendance after completing training (see Figure 7C).

**Differentiating training condition effects from effects of other variables.**

We considered the possibility that the positive pre-post change in working memory, primary drug substance use frequency, and TAU session attendance displayed by participants in the CogTrain condition could reflect influences from variables outside of training condition. We examined whether additional covariates were needed in subsequent analyses based on individual differences in attendance at the posttreatment session and baseline differences between participants assigned to different training conditions. Attendance at the posttreatment session was related to pretreatment trait anxiety (TAI), $r_s(44) = .346$, $p = .019$, perceived stress (PSS), $r_s(44) = .295$, $p = .047$, and change and motivation readiness (CMR), $r_s(44) = -.302$, $p = .049$, with individuals who reported higher levels of trait anxiety, perceived stress, and levels of change and motivation readiness being more likely to present at posttreatment. We included these variables, as well as pretreatment distress tolerance, which differed between training conditions, in the next series of analyses.

To examine whether targeted training resulted in improvements in these cognitive and real-world variables when accounting for variables that differed between individuals assigned to different training conditions (i.e., baseline distress tolerance) and participants who did versus did not attend posttreatment (i.e., pretreatment trait anxiety, perceived stress, change and motivation readiness), we re-ran repeated measures ANOVAs including each of these variables as a continuous covariate in separate analyses.

For analyses of working memory and primary drug substance use frequency, the timepoint by condition effect remained significant when controlling for the influence of each of these potential confounding variables [distress tolerance on working memory:

$F(1, 28) = 6.12$, $p = .020$, $\eta^2 = .18$; trait anxiety on working memory: $F(1, 28) = 8.31$, $p = .007$, $\eta^2 = .23$; perceived stress on working memory: $F(1, 28) = 8.06$, $p = .008$, $\eta^2 = .22$, change and motivation readiness on working memory: $F(1, 26) = 9.57$, $p = .005$, $\eta^2 = .27$], distress tolerance on substance use: $F(1, 30) = 7.99$, $p = .008$, $\eta^2 = .21$; trait anxiety on substance use: $F(1, 30) = 7.19$, $p = .012$, $\eta^2 = .19$; perceived stress on substance use: $F(1, 30) = 7.17$, $p = .012$, $\eta^2 = .19$; change and motivation readiness on substance use: $F(1, 28) = 5.55$, $p = .026$, $\eta^2 = .17$)]. The timepoint by condition effect also remained for TAU session attendance when controlling for trait anxiety, $F(1, 35) = 4.97$, $p = .032$, $\eta^2 = .12$, perceived stress, $F(1, 35) = 4.80$, $p = .035$, $\eta^2 = .12$; and change and motivation readiness, $F(1, 33) = 5.11$, $p = .030$, $\eta^2 = .13$. These results suggest between condition differences in pre-post change are reflective of training condition effects, rather than alternative variables that were not experimentally manipulated.

However, the timepoint by training condition interaction for TAU session attendance was no longer significant when controlling for variability in distress tolerance among participants, $F(1, 35) = 2.00$, $p = .166$, $\eta^2 = .05$. Instead, when controlling for the influence of distress tolerance, the main effect of timepoint on TAU session attendance was significant, $F(1, 35) = 4.34$, $p = .045$, $\eta^2 = .11$, such that individuals in both training conditions attended more sessions at posttreatment compared to pretreatment. The interaction between timepoint and distress tolerance indicated a trend, $F(1, 35) = 3.85$, $p = .058$, $\eta^2 = .10$, with a tendency for individuals with lower baseline distress tolerance to display greater increases in TAU session attendance from pretreatment to posttreatment, compared to individuals with higher baseline distress tolerance. This series of results raised the possibility that the observed change in TAU attendance among individuals in

the CogTrain condition might be accounted for by the low levels of distress tolerance they exhibited at a group level, and the tendency for individuals with low distress tolerance at pretreatment to show increases in session attendance between pretreatment and posttreatment. However, results for this analysis using multiple imputation for missing posttreatment values, suggest this finding is the result of reduced statistical power, since the two-way interaction between timepoint and training condition remained significant when controlling for the influence of pretreatment distress tolerance and including imputed data, B = 1.05, SE = 0.486, 95% CI = 0.130, 1.977, t = 2.28, $p$ = 0.028; see Supplementary Material for more detail).

***Pre-post change as a function of externalizing traits***

To examine whether pre-post change in working memory (Digit Backwards) and real-world behavior (primary substance use frequency and TAU session attendance) varied as a function of externalizing traits, we included ESI total score as a covariate in each of the repeated measures ANOVAs. Controlling for externalizing, the interaction between timepoint and condition remained significant for working memory, $F(1, 28) = 7.84$, $p = .009$, $\eta^2 = .22$, primary drug substance use frequency, $F(1, 30) = 10.52$, $p = .003$, $\eta^2 = .26$, and TAU session attendance, $F(1, 35) = 4.85$, $p = .034$, $\eta^2 = .12$. Moreover, for the analyses of working memory and TAU session attendance, neither the main effect of externalizing, nor the interaction between timepoint and externalizing, were significant (all *p's* > .338*).* However, for the analysis of primary drug substance use frequency, there was a significant interaction between timepoint and level of trait externalizing, $F(1, 30) = 10.52$, $p = .003$, $\eta^2 = .26$, such that higher levels of externalizing were associated with increases in primary drug substance use from pretreatment to

posttreatment and lower levels of externalizing were associated with decreases in primary drug substance use over the same period. This finding is consistent with previous studies indicating poor treatment outcomes among antisocial individuals, particularly those high on trait externalizing, in substance use treatment (Compton et al., 2003; Thylstrup & Hesse, 2016).

To examine whether externalizing traits also moderated the interaction between timepoint and training condition on substance use, we ran an additional repeated measures ANOVA including the three-way interaction among timepoint, training condition, and ESI-Brief total scores. In this model, the three-way interaction was not significant, $F(1, 29) = 2.82$, $p = .104$, $\eta^2 = .09$, while the two-way interaction between timepoint and condition and the two-way interaction between timepoint and externalizing remained significant (timepoint-by-condition: $F(1, 29) = 10.45$, $p = .003$, $\eta^2 = .27$; timepoint-by-externalizing: $F(1, 29) = 11.08$, $p = .002$, $\eta^2 = .28$). On the whole, this pattern of results indicated that externalizing was associated with diminished substance use reductions over the course of treatment. However, despite the overall tendency for externalizing to limit positive pre-post change, individuals who completed CogTrain, the targeted cognitive remediation training, demonstrated greater reductions in substance use frequency compared to those who completed control training, regardless of their level of externalizing.

***Pre-post change by training dose***

Attendance at training sessions was variable, with individuals completing between one and eight total training sessions. Of note, for individuals who attended the posttreatment session, there was less variability in training session attendance, with 89%

of those who presented at posttreatment attending five or more sessions. For those who did not attend the posttreatment session, no participants attended more than four training sessions. It is nonetheless possible that differential exposure to training (i.e., completing all versus fewer training sessions) influenced the observed relationships among training condition, timepoint, and changes in cognitive performance and real-world behavior. To explore this possibility, we re-ran the initial repeated measures ANOVAs including the main effect of the number of training sessions, and the two- and three-way interactions with timepoint and condition as additional continuous fixed effect predictors of working memory, primary drug substance use frequency, and TAU session attendance. Across analyses examining training exposure, pre-post change associated with the CogTrain condition did not appear to be influenced by training session dose, with none of the three-way interactions among training dose, timepoint, and condition being significant (all *p*'s > .103). Thus, CogTrain appears to be effective, at minimum, for individuals who attended five or more training sessions.

***Pre-post change across measures by training condition***

To determine whether participants who showed improvements with CogTrain training improved across multiple domains, rather than on just one outcome measure (e.g., working memory but not substance use frequency or TAU session attendance), we examined whether the significant improvement observed across three outcome measures reflected improvement on a single factor, and whether training condition influenced overall change. We first computed difference scores quantifying improvement on each measure (i.e., increased working memory capacity, decreased primary drug substance use frequency, increased TAU session attendance) and entered them into a PCA with

orthogonal rotation (varimax). This analysis indicated a one-factor solution, with the single factor representing improvement across all three outcome measures ($r_{\text{working memory}}(26) = 0.86$, $p < .001$, $r_{\text{substance use}}(26) = 0.51$, $p = .006$, $r_{TAU\ \text{attendance}}(26) = 0.77$, $p < .001$; see Table 4 for descriptive statistics).

Next, we entered the factor score into an ANOVA, with training condition (0 = Active Control, 1 = CogTrain) as a categorical predictor of the pre-post improvement factor score. Consistent with the results of the separate repeated measures ANOVAs, there was a significant relationship between training condition and pre-post change across measures, $F(1, 26) = 10.66$, $p = .003$, $R^2 = .54$, with membership in the CogTrain condition (compared to the control condition) predicting positive pre-post change across outcome measures (B = 1.03, SE = 0.31, 95% CI = 0.38, 1.67, $t = 3.27$, $p = .003$). Results of these analyses suggest that individuals who improved on working memory, primary drug substance use frequency, and TAU session attendance did so across all three measures, and that improvement across measures was indeed associated with completing the targeted cognitive remediation training.

***The impact of improvement on training on pre-post change***

Finally, we examined whether improvement on any specific CogTrain task predicted overall pre-post change in cognitive functioning and real-world behavior. We analyzed the overall pre-post change factor score for each participant in a General Linear Model (GLM) with Beta values for each CogTrain task variable (neutral and rewarded SSRTs, Stop Signal Task reward scores, percentage of ambiguous choices, persistence; z-scored) as continuous predictors (see Table 4 for descriptive statistics and zero-order correlations among the factor associated with pre-post across outcome measures and

training task Beta values). There was a significant relationship between training Beta's and pre-post change across measures ($p$ = .030). Improvement on inhibition for both neutral and rewarded trials (having lower SSRT's across sessions) and on persistence during the PASAT-C (higher persistence scores across sessions) was associated with more positive pre-post change across outcome measures (SSRT for neutral trials: B = 0.57, SE = 0.19, 95% CI = 0.19, 0.95, $p$ = .003; SSRT for reward trials: B = 0.50, SE = 0.16, 95% CI = 0.18, 0.81, $p$ = .002; PASAT-C persistence: B = 0.55, SE = 0.16, 95% CI = 0.23, 0.88, $p$ = .001). Improvement on Stop Signal Task scores for rewarded trials (higher scores across sessions) and improvement on ambiguous choices during decision making (fewer ambiguous choices) were not significantly associated with pre-post change across outcome measures (Stop Signal Task reward scores: B = -0.15, SE = 0.18, 95% CI = -0.50, 0.20, $p$ = .402; ambiguous choices: B = -0.31, SE = 0.19, 95% CI = -0.68, 0.05, $p$ = .094).

When we ran a separate GLM for the control training condition, entering the Beta value for Text Twist word scores (z-scored) as a predictor of pre-post change across outcome measures, there was no significant relationship between control training change and change on cognitive performance and real-world behavior, B = -0.19, SE = 0.17, 95% CI = -0.51, 0.14, $p$ = .265. Overall, these results suggest that improvement on the CogTrain tasks, and not the control training, was associated with generalized improvement on outcome measures. Improving inhibition in neutral and rewarding contexts and persistence under distress may be particularly conducive to enhanced working memory, decreased substance use, and increased engagement in substance use

treatment.

## Discussion

Previous research highlights how several cognitive-affective dysfunctions relate to the chronic maladaptive behaviors displayed by antisocial individuals. The results of the present study suggest that targeting these dysfunctions may not only mitigate such aberrations, but also may lead to changes in more general cognitive functioning and antisocial behaviors themselves. Through the use of a novel cognitive remediation training package designed to enhance inhibition in the presence of reward, decision-making under ambiguity, and working memory amid distress, this pilot randomized clinical trial provides evidence to support the longstanding but largely untested claims that treatment adjuncts or adaptations addressing cognitive-affective dysfunctions are likely to improve treatment efficacy among antisocial individuals (Ball et al., 2006; Daughters et al., 2008; Wölwer et al., 2001). Moreover, by linking improvements in cognitive-affective performance with improvements in real-world behavior, this study represents a substantial step in bridging the gap between behavior, mechanism, and treatment in a population that has historically been viewed as treatment-resistant (Raine, 2018).

As predicted, antisocial individuals in outpatient substance use treatment who completed targeted training tasks matched to their cognitive-affective dysfunctions demonstrated improved performance on training tasks with repeated practice. Specifically, they became more efficient and more accurate when inhibiting prepotent responses in neutral and rewarding contexts, respectively. They also were more

considerate of ambiguous information about costs and benefits when making decisions and more persistent when demands were placed on working memory amidst distress. More importantly, individuals who completed CogTrain training demonstrated improvements outside of trained tasks on a pre-post assessment battery, with greater working memory capacity, reduced substance use frequency for their primary drug of choice, and better treatment session attendance. Reflecting a relationship between training improvement and the observed generalized changes in cognition and behavior, overall improvements in working memory, substance use, and TAU session attendance correlated with improvements on CogTrain tasks. Additionally, observed training-related improvements in cognitive and behavioral functioning occurred regardless of levels of externalizing. Results were strengthened by our inclusion of an active control training group for comparison. Although individuals who completed the active control training (a computerized word game) improved on the task they practiced, they did not show the improvements in general cognitive functioning or real-world behavior displayed by individuals in the CogTrain condition. Overall, training that targeted the cognitive-affective dysfunctions associated with antisociality successfully improved functioning in intended domains and led to more global improvements in functioning.

**Considerations for the CogTrain Training Package**

Results from the present study replicate and expand upon findings from prior studies utilizing cognitive remediation training in psychiatric populations, including the only other previous study examining cognitive remediation in antisocial individuals (Baskin-Sommers et al., 2015), and several studies examining cognitive remediation in adults with other externalizing disorders (i.e., attention deficity hyperactivity disorder;

Dentz, Guay, Parent, & Romo, 2020; substance use disorders; Verdejo-Garcia, 2016). Across studies of externalizing individuals more broadly, it appears that both response inhibition (Baskin-Sommers et al., 2015; Houben, Havermans, Nederkoorn, & Jansen, 2012; Houben, Nederkoorn, Wiers, & Jansen, 2011) and working memory deficits (Bickel et al., 2011; Gamito et al., 2014; Houben, Wiers, & Jansen, 2011; Rass et al., 2015; Rupp et al., 2012) can be improved upon with cognitive remediation training targeting these domains. However, the extent to which such training results in transfer to other tasks and clinical outcomes is not consistent in prior studies of externalizing individuals (Baskin-Sommers et al., 2015; Bickel et al., 2011; Gamito et al., 2014; Houben, Wiers, et al., 2011; Rass et al., 2015; Rupp et al., 2012; Verdejo-Garcia, 2016) and other psychiatric populations (Kim et al., 2018). The generalized improvements in antisocial individuals demonstrated in the present study, along with their medium to large effect sizes, likely are attributable to specific characteristics of the CogTrain cognitive remediation training package.

First, our training package included three tasks spanning several cognitive functions, rather than just one specific function or area of cognition. While antisociality previously has been related to problems in executive functioning more broadly, considerable empirical evidence suggests this characterization is oversimplified, with dysfunction being specific to some discrete executive functions (e.g., working memory, inhibition; Chamberlain et al., 2016; De Brito et al., 2013; Dolan & Park, 2002; Rubio, Jimenez, et al., 2007; Swann et al., 2009; Zeier et al., 2012) and not others (e.g., planning, set-shifting; Chamberlain et al., 2016; Dolan & Park, 2002; Stevens et al., 2003), and in cognitive domains outside of executive functioning (e.g. perception, Bauer,

2001; attention, Lijffijt et al., 2009; 2012, decision-making, Buckholtz et al., 2017; Mazas et al., 2000; Petry, 2002; Swogger, Walsh, Lejuez, & Kosson, 2010). The complex and multifaceted nature of dysfunction in antisocial individuals requires treatment that is similarly multifaceted; treatments targeting a single cognitive domain in isolation are unlikely to impact overall functioning (Zalta & Shankman, 2016).

Second, our study was unique in its deliberate effort to target functioning at the intersection of both cognition *and* affect to increase generalizability to real-world behavior. Although the training tasks designed by Baskin-Sommers and colleagues (2015) purportedly centered around "affective-cognitive control," only one of the three tasks (incentivized Go-Stop; Albrecht, Banaschewski, Brandeis, Heinrich, & Rothenberger, 2005; Avila & Parcet, 2001; Schuckit et al., 2012) targeted both cognition and affect simultaneously, while the other two tasks focused on cognition or affect alone. By contrast, all three CogTrain tasks placed dual demands on cognition and affect, increasing the likelihood that our training package prepared individuals to practice skills in contexts that more closely matched the environments typical of real-life situations in which they might struggle—environments characterized by strong cravings, unknown risks, and distress. Indeed, performance measures from the training tasks selected previously have been linked to relevant real-world outcomes for antisocial individuals, such as arrest frequency, substance use relapse, and treatment dropout (Buckholtz et al., 2017; Daughters, Lejuez, Bornovalova, et al., 2005; Konova et al., 2020). Emotional cues are important triggers of maladaptive behavior in antisocial individuals, and emotional exposure is likely necessary to provide adequate practice replacing problematic behaviors with more adaptive responses (Bornovalova, Gratz, Daughters, Hunt, & Lejuez, 2012;

Brown et al., 2008). Antisocial individuals lack both the "cognitive and affective equipment required" to follow rules and respect the rights of others (Alcázar-Córcoles, Verdejo-García, Bouso-Saiz, & Bezos-Saldana, 2010, p. 291). Future efforts to change behaviors among antisocial individuals will benefit from not only bolstering their cognitive capacities, but also by doing so in environments that are affectively-charged and thus more ecologically valid.

Despite the noted associations with positive pre-post change, the CogTrain package could benefit from refinement. For the Stop Signal Task, inhibition reward scores for rewarded trials did not improve significantly with practice (the effect of session indicated a statistical trend), and the reward score performance metric, which did improve significantly, was not associated with improvements on pre-post tasks. Recent research examining the impact of incentives on inhibitory performance using the Stop Signal Task highlights complex interactions among expectation, magnitude, and order of rewards among healthy individuals (Herrera et al., 2019). Future research examining the impact of specific reward manipulations on inhibition among antisocial individuals may further elucidate how specific reward features disrupt antisocial individuals, and how to address these disruptions. Alternatively, other response inhibition tasks with more established and less complex reward effects among antisocial individuals (e.g., modified *n*-back; Baskin-Sommers et al. 2014; Stuppy-Sullivan & Baskin-Sommers, 2019) may be used in future training packages.

For the decision-making task, there was no significant training change in the ambiguity attitude parameter (β). This finding is consistent with recently published work utilizing a similar decision-making task in individuals with opioid disorders, who showed

higher variability in ambiguity attitude across sessions compared to healthy controls (Konova et al., 2020). Antisocial individuals may be similarly unlikely to provide reliable ambiguity attitudes as measured by β. Notably, the decision-making task yielded more excluded trials than other tasks in the training package, due to several participants selecting the lottery option on multiple catch trials and computational models providing poor fit to data for several individuals' sessions (see Supplementary Material). By contrast, the percentage of ambiguous choices variable did show significant improvement across training sessions. This pattern of findings aligns with our prior findings from Study 2 suggesting antisocial individuals have difficulty integrating multiple decision variables when making choices, and that their behavior may be better explained by simple computational models with fewer parameters (i.e., reflecting a general bias toward ambiguous versus certain options). Given their documented difficulty with complex decisions and tasks, future efforts to encourage antisocial individuals to consider ambiguous information during decision-making may benefit from introducing decision parameters slowly (e.g., starting with variability in reward magnitude and known probability, and later adding varying levels of ambiguity, effort, etc.). In general, antisocial individuals also may benefit from being given more clear instructions, additional incentives for paying attention, and other strategies that ensure they are engaging in tasks optimally.

When evaluating the CogTrain package as a whole, it is important to consider that attendance at CogTrain training sessions was lower than attendance at control training sessions. It is possible that lower attendance among individuals in the CogTrain condition was related to their lower baseline distress tolerance or treatment attendance; however, it

also is possible that the training itself contributed to between-condition discrepancies in training session attendance. For example, prior studies examining the Stop Signal Task and the PASAT-C indicate that both tasks produce strong negative affect, including frustration, anxiety, and discomfort (Bornovalova et al., 2012; Gratz, Bornovalova, Delany-Brumsey, Nick, & Lejuez, 2007; Spunt, Lieberman, Cohen, & Eisenberger, 2012). Studies in nonclinical samples suggest that negative affect prompted after deploying cognitive control impacts decisions about approaching or avoiding cognitively demanding tasks in the future (Kool, McGuire, Rosen, & Botvinick, 2010; Spunt et al., 2012). Thus, discomfort induced while practicing inhibition in an environment designed to produce errors (i.e., using the dual-staircase tracking procedure on the Stop Signal Task) and while challenging working memory amid distress (i.e., created by accelerating the pace of subsequent stimuli and administering noise blasts during the PASAT-C) may have decreased willingness to attend future study sessions among individuals in the participants receiving CogTrain. In turn, perhaps only the individuals with sufficient ability to tolerate such discomfort completed sufficient training sessions, gleaned generalized benefits, and presented at posttreatment. While we do not have measures of subjective experiences of training tasks to confirm connections between negative affect and training session attendance in the CogTrain condition, our results raise the possibility that adherence to adjunctive treatment protocols may benefit from additional accommodations for antisocial individuals based on their cognitive-affective experience. For example, concluding treatment sessions with relaxation exercises (a common component of exposure-based treatments), and adjusting training tasks to accommodate a wider range of difficulty levels, may decrease the extent to which negative affect prompts

treatment disengagement. Ultimately, changing maladaptive behavior likely requires sensitivity not only to the cognitive-affective dysfunctions typical of a given psychiatric population, but also to individual differences within such a population.

**Translation to Cognitive-Affective Functioning**

As noted above, there was evidence of generalized change for working memory, primary drug substance use frequency, and TAU session attendance. However, the CogTrain training package did not lead to significant pre-post change across all measures of the cognitive-affective battery. The non-significant trend for pre-post by condition change in distress tolerance may relate to the poor convergence between measures of behavioral distress tolerance found in previous studies (Glassman et al., 2016; McHugh et al., 2011). The breath holding task used to measure pre-post change is considered to be a measure of somatic persistence, and induces affective changes that are more fleeting than those induced by the PASAT-C, a cognitive measure of distress tolerance (McHugh et al., 2011). Although we deliberately selected the breath holding task as part of a conservative effort to select broad-based measures of change, this choice may have restricted our ability to detect transfer effects. Selecting another measure of distress tolerance with more cognitive-affective focus (e.g., the computerized Mirror-Tracing Persistence Task; MTPT-C; Strong et al., 2003) may have produced more apparent transfer effects. Another possible hindrance to detecting pre-post change in distress tolerance was the between-condition difference in distress tolerance observed at baseline and our small sample size. With unequal groups at the start of training, and variability in distress tolerance across all participants, potential CogTrain-related within-subject changes in distress tolerance may have been subdued and, with limited power to observe small effects, reduced to a trend.

Thus, future studies with larger sample sizes and an alternative selection of measures are needed to determine whether training working memory under distress can positively impact distress tolerance.

Delay discounting also was not impacted by the CogTrain package, representing a divergence from previous cognitive remediation studies that suggest a connection between working memory training and reductions in delay discounting (Bickel et al., 2011; Felton, Collado, Ingram, Doran, & Yi, 2019). Null effects observed in the present study may have occurred for a number of methodological and theoretical reasons. Methodologically, our 5-item measure of delay discounting was briefer, and possibly less precise, than the longer adjusting amount procedures utilized in previous studies, reducing our ability to detect meaningful change in preference for immediate rewards. Moreover, the magnitude of our delayed reward ($100) was relatively small. Felton and colleagues (2019) found a connection between working memory change and delay discounting change under a $1000 delayed reward, but not for a $50 delayed reward. Recent evidence suggests that within- and between-group differences in delay discounting are more likely to be revealed in larger magnitude conditions (Felton et al., 2019; Mellis, Woodford, Stein, & Bickel, 2017). On a more conceptual note, while the PASAT-C task in the CogTrain package placed demands on working memory, difficulty was progressively increased by increasing the rate of stimulus presentation and increasing distress through use of a noise blast, not by increasing set sizes, as in working memory training tasks used in other studies (e.g., PSSCogReHab; Psychological Software Services, Indianapolis, Indiana). It is possible that the task employed in the present study tapped an overlapping but distinct working memory application from previous studies,

with differential links to delay discounting. Overall, null delay discounting results further highlight the importance of specificity when characterizing cognitive-affective dysfunctions (e.g., applications of working memory, types of cost-benefit decision-making) relevant to antisocial individuals and the factors (e.g.., reward magnitude, relations to other cognitive-affective domains) that influence them.

**Limitations**

Beyond specific considerations for the CogTrain package and measures of generalized change, it is important to note some additional limitations. First, and perhaps most notably, are limitations related to our premature ending to data collection and small subsequent sample size. Recruiting fewer participants than our target sample size may have biased randomization, resulting in the between-condition imbalances in baseline TAU session attendance and distress tolerance. This is because temporal trends in patient recruitment are common in randomized controlled trials using urn randomization (Friedman et al., 2015; Wayant, Tritz, Horn, Crow, and Vassar, 2020). At the time data collection ended (roughly halfway to our target *N),* individuals in the CogTrain condition differed from individuals in the Control condition on two urn variables: they had been in treatment for shorter durations and had used their primary drugs on more days in the past month as of their pretreatment assessments. While speculative, these differences could correlate with lower levels of distress tolerance and lower TAU session attendance at baseline among individuals in the CogTrain group. Although a majority of our results remained significant when controlling for potential covariates and when imputing missing data (see Supplementary Material), a fully powered study would have been better equipped to determine which domains of functioning are likely to be improved by

cognitive remediation training, and which individual differences may moderate improvements for antisocial individuals, with greater confidence.

Another, and potentially related, limitation of the present study was our inability to detect possible dose effects due to a lack of variability in attendance among individuals who presented at posttreatment. Our results suggest that attendance at more than half of the required training sessions was linked with training-related gains. It will be important for future cognitive remediation studies in this population to collect large enough samples to examine training dose effects fully. While it appears that single-session training can impact real-world outcomes (e.g., substance use; Verdejo Garcia, Houben 2011, Houben 2012), it is unclear if abbreviated cognitive remediation protocols can produce change in individuals with clinically significant levels of antisociality. Future studies might also consider adjusting the number of training sessions based on individual differences in training task performance (e.g., Bickel et al., 2011), since providing no more and no fewer training sessions than necessary will ensure such training is as cost-effective as possible.

Finally, while we measured multiple real-world behaviors relevant to antisocial individuals (i.e., substance use frequency, treatment attendance), we did not measure many that are typical of the population (e.g., engagement in crime, acts of aggression, risky sexual behavior) and reflect important outcomes relevant to the impact of their behavior. Future treatment studies evaluating pre-post change in behavior will benefit from assessing engagement in a more diverse range of antisocial behaviors and denoting which behaviors are the most persistent as targets for future empirical studies and treatments.

## Conclusions

In sum, results from the present study provide evidence that antisocial individuals can learn to change their behavior, despite the historic pessimism surrounding their potential to respond to treatment. Importantly, behavior change among antisocial individuals is made possible by more richly understanding the complex cognition-emotion interactions impacting their functioning and properly leveraging such knowledge into targeted treatments. Toward this end, computerized cognitive remediation training is an exciting potential tool. With widespread access to Internet-enabled mobile technology (Pew Research Center, 2018), cognitive remediation may not only help bridge the science-practice gap (Onken, Carroll, Shoham, Cuthbert, & Riddle, 2014), but also the gap between mechanism-informed interventions and the individuals for whom they are intended, supporting behavior change in a wider variety of settings (e.g., primary care, rural areas), and at lower costs than traditional behavior change interventions.

## Tables: Study 3

### Table 1

*Baseline demographics, psychiatric disorders, and assessment measures by training condition.*

| Variable | Cognitive Remediation (N = 22) | | Active Control (N = 24) | | Total (N = 46) | | Analysis | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | *%* | *N* | *%* | *N* | *%* | *χ2* | *df* | *p* |
| Female | 11 | 50.0 | 11 | 45.8 | 22 | 47.8 | 0.080 | 1 | .777 |
| Race | | | | | | | | | |
| White | 12 | 54.5 | 17 | 70.8 | 29 | 67.4 | 0.943 | 1 | .331 |
| Black | 8 | 36.4 | 6 | 25.0 | 11 | 23.9 | | | |
| Unreported | 2 | 9.1 | 1 | 4.2 | 6 | 13.0 | | | |
| Hispanic Ethnicity | 7 | 31.8 | 5 | 20.8 | 12 | 26.1 | 0.718 | 1 | .397 |
| In treatment 30 days or more | 6 | 27.3 | 11 | 45.8 | 17 | 37.0 | 1.697 | 1 | .193 |
| Single or Divorced | 18 | 81.8 | 19 | 79.2 | 37 | 80.4 | 0.051 | 1 | .821 |
| Unemployed | 8 | 36.4 | 13 | 54.2 | 21 | 45.7 | 1.466 | 1 | .226 |
| Obtained a high school diploma or GED | 21 | 95.5 | 22 | 91.7 | 43 | 93.5 | 0.270 | 1 | .603 |
| Primary Substance Use Disorder (SUD) | | | | | | | | | |
| Alcohol | 7 | 31.8 | 8 | 33.3 | 15 | 32.6 | 2.337 | 3 | .506 |
| Opioids | 7 | 31.8 | 8 | 33.3 | 15 | 32.6 | | | |
| Marijuana | 3 | 13.6 | 6 | 25.0 | 9 | 19.6 | | | |
| Cocaine | 5 | 22.7 | 2 | 8.3 | 7 | 15.2 | | | |
| Primary SUD symptom severity | | | | | | | | | |
| Mild (2-3 symptoms) | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1.843 | 1 | .175 |
| Moderate (4-5 symptoms) | 5 | 22.7 | 2 | 8.3 | 7 | 15.2 | | | |
| Severe (6+ symptoms) | 17 | 77.3 | 22 | 91.7 | 39 | 84.8 | | | |
| Total number of lifetime SUDs | | | | | | | | | |
| 1 | 4 | 18.2 | 3 | 12.5 | 7 | 15.2 | 0.657 | 2 | .720 |
| 2 to 3 | 12 | 54.5 | 12 | 50.0 | 24 | 52.2 | | | |
| 4 or more | 6 | 27.3 | 9 | 37.5 | 15 | 32.6 | | | |
| Lifetime Alcohol Use Disorder | 18 | 81.8 | 16 | 66.7 | 34 | 73.9 | 1.367 | 1 | .242 |
| Lifetime Stimulant Use Disorder | 15 | 68.2 | 14 | 58.3 | 29 | 63.0 | 0.478 | 1 | .489 |
| Lifetime Marijuana Use Disorder | 14 | 63.6 | 16 | 66.7 | 30 | 65.2 | 0.046 | 1 | .829 |
| Lifetime Opioid Use Disorder | 10 | 45.5 | 11 | 45.8 | 21 | 45.6 | 0.001 | 1 | .979 |
| Antisocial Personality Disorder | 1 | 0.05 | 1 | 0.04 | 2 | 0,04 | 0.009 | 1 | .923 |
| Any lifetime depressive disorder | 11 | 50.0 | 17 | 70.8 | 28 | 60.9 | 2.092 | 1 | .148 |
| Any lifetime anxiety disorder | 11 | 50.0 | 11 | 45.8 | 22 | 47.8 | 0.080 | 1 | .777 |
| Lifetime Posttraumatic Stress Disorder | 10 | 45.5 | 9 | 37.5 | 19 | 41.3 | 0.300 | 1 | .584 |
| | *M* | *SD* | *M* | *SD* | *Mean* | *SD* | *F* | *df* | *p* |
| Age | 34.9 | 8.1 | 32.4 | 8.6 | 33.6 | 8.4 | 1.017 | (1, 44) | .319 |
| Years primary substance used | 13.9 | 8.4 | 12.1 | 7.7 | 13.0 | 8.0 | 0.591 | (1, 44) | .446 |
| Full Scale IQ (Shipley) | 100.0 | 14.0 | 101.5 | 13.3 | 100.8 | 13.5 | 0.132 | (1, 44) | .718 |
| Number of attended training sessions | 5.5 | 2.6 | 6.9 | 1.8 | 6.2 | 2.3 | 4.716 | (1, 44) | .035* |
| Pretreatment Assessments | | | | | | | | | |
| Externalizing Spectrum Inventory-Brief Total Score | 223.1 | 52.1 | 210.3 | 51.8 | 216.4 | 51.8 | 0.697 | (1, 44) | .408 |
| State Anxiety Score | 37.2 | 14.0 | 41.1 | 11.2 | 39.3 | 12.6 | 1.099 | (1, 44) | .300 |
| Trait Anxiety Score | 42.4 | 12.5 | 47.5 | 9.8 | 45.1 | 11.3 | 2.422 | (1, 44) | .127 |
| Perceived Stress Total Score | 18.5 | 6.7 | 21.4 | 4.6 | 20.0 | 5.8 | 2.929 | (1, 44) | .094 |
| Change and Motivation Readiness Score | 58.4 | 10.1 | 60.0 | 10.5 | 59.0 | 10.2 | 0.148 | (1, 44) | .703 |
| Cognitive Affective Battery | | | | | | | | | |
| Digit Span Backwards Score | 7.0 | 2.1 | 7.4 | 2.9 | 7.2 | 2.5 | 0.248 | (1, 37) | .622 |
| Distress Tolerance Score | 2.2 | 0.8 | 2.7 | 0.7 | 2.4 | 0.8 | 5.365 | (1, 44) | .025* |
| Ln(*k)* | -2.0 | 1.8 | -2.3 | 2.1 | -2.18 | 2.0 | 0.241 | (1, 44) | .626 |
| Real World Behavior | | | | | | | | | |
| Days Primary Substance Use (cube-rooted) | 1.0 | 1.1 | 0.6 | 1.0 | 0.8 | 1.0 | 1.047 | (1.44) | 0.312 |
| TAU Session Attendance | 4.4 | 3.2 | 8.2 | 6.6 | 6.4 | 5.5 | 6.205 | (1.44) | 0.017* |

**$p$<0.01, *$p$>.0.05

**Table 2**

*Training and study session attendance by training condition.*

| | Cognitive Remediation (N = 22) | | Active Control (N = 24) | | Total (N = 46) | | Analysis | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | N | % | N | % | N | % | | | |
| Total Training Sessions Completed | | | | | | | | | |
| 1 | 2 | 9.1 | 0 | 0.0 | 2 | 4.3 | | | |
| 2 | 2 | 9.1 | 1 | 4.2 | 3 | 6.5 | | | |
| 3 | 2 | 9.1 | 1 | 4.2 | 3 | 6.5 | | | |
| 4 | 2 | 9.1 | 2 | 8.3 | 4 | 8.7 | | | |
| 5 | 2 | 9.1 | 1 | 4.2 | 3 | 6.5 | | | |
| 6 | 2 | 9.1 | 1 | 4.2 | 3 | 6.5 | | | |
| 7 | 2 | 9.1 | 3 | 12.5 | 5 | 10.9 | | | |
| 8 | 8 | 36.4 | 15 | 62.5 | 23 | 50.0 | | | |
| | M | SD | M | SD | M | SD | F | df | *p* |
| Total Training Sessions Completed | 5.5 | 2.6 | 6.9 | 1.8 | 6.2 | 2.3 | 4.716 | 1,44 | .035* |
| | N | % | N | % | N | % | $\chi 2$ | df | *p* |
| Attended Posttreatment Session | 15 | 68.2 | 23 | 95.8 | 38 | 82.6 | 6.109 | 1 | .013* |
| Attended follow-up session | 12 | 54.5 | 20 | 83.3 | 32 | 69.6 | 4.493 | 1 | .034* |

**$p$<0.01, *$p$>.0.05

**Table 3**

*Descriptive statistics and zero-order correlations for individual differences and training task variables.*

| | | | | *r* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | *n* | *M* | *SD* | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| 1. ESI Total Score | 46 | 216.4 | 51.8 | 1.000 | | | | | | | | | |
| 2. Full Scale IQ | 46 | 100.8 | 13.5 | 0.327* | 1.000 | | | | | | | | |
| 3. SSRT neutral blocks | 19 | 571.8 | 116.4 | -0.261 | -0.001 | 1.000 | | | | | | | |
| 4. SSRT reward blocks | 17 | 599.3 | 39.2 | 0.077 | 0.267 | 0.422 | 1.000 | | | | | | |
| 5. Stop Signal Task reward score | 15 | 532.2 | 20.2 | 0.144 | 0.427 | 0.547* | 0.141 | 1.000 | | | | | |
| 6. % Risky Choices | 22 | 0.70 | 0.17 | 0.434* | -0.152 | -0.293 | -0.256 | -0.148 | 1.000 | | | | |
| 7. % Ambiguous Choices | 22 | 0.56 | 0.22 | 0.223 | -0.518* | -0.077 | -0.275 | -0.206 | 0.375 | 1.000 | | | |
| 8. Risk Attitude (α) | 21 | 0.85 | 0.77 | 0.024 | 0.473* | -0.178 | 0.040 | 0.056 | -0.167 | -0.603** | 1.000 | | |
| 9. Ambiguity Attitude (β) | 22 | 0.58 | 0.93 | -0.040 | 0.008 | -0.044 | -0.171 | -0.163 | 0.020 | -0.308 | 0.408 | 1.000 | |
| 10. PASAT-C Persistence | 21 | 0.39 | 0.24 | 0.223 | 0.464* | 0.330 | 0.136 | 0.608* | -0.002 | -0.315 | -0.272 | 0.204 | 1.000 |
| 11. Text Twist Words Completed | 24 | 136.4 | 56.0 | -0.179 | 0.598** | -- | -- | -- | -- | -- | -- | -- | -- |

**$p<0.01$, *$p>.0.05$

**Table 4**

*Descriptive statistics and zero-order correlations for overall pre-post change and across-session change on training tasks.*

| | | | | *r* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | *n* | *M* | *SD* | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 1. Pre-Post Change Factor Score | 28 | 0.135 | 0.967 | 1.000 | | | | | | |
| 2. Beta SSRT neutral | 18 | 0.014 | 0.022 | 0.139 | 1.000 | | | | | |
| 3. Beta SSRT reward | 18 | 0.017 | 0.033 | 0.219 | 0.212 | 1.000 | | | | |
| 4. Beta Stop Signal Task reward score | 18 | -0.003 | 0.062 | -0.530 | 0.171 | 0.018 | 1.000 | | | |
| 5. Beta % Ambiguous Choices | 18 | -2.58 | 13.87 | 0.174 | -0.115 | -0.506* | -0.200 | 1.000 | | |
| 6. Beta PASAT-C Persistence | 15 | 13.36 | 7.12 | -0.403 | -0.258 | -0.059 | -0.409 | 0.130 | 1.00 | |
| 7. Beta Text Twist | 24 | 0.058 | 0.027 | -0.277 | -- | -- | -- | -- | -- | 1.00 |

**$p$<0.01, *$p$>.0.05

**References: Study 3**

Albrecht, B., Banaschewski, T., Brandeis, D., Heinrich, H., & Rothenberger, A. (2005). Response inhibition deficits in externalizing child psychiatric disorders: An ERP-study with the Stop-task. *Behavioral and Brain Functions, 1*(1), 1-14.

Alcázar-Córcoles, M. Á., Verdejo-García, A., Bouso-Saiz, J. C., & Bezos-Saldana, L. (2010). Neuropsychology of impulsive aggression. *Revista de Neurología, 50*(5), 291-299.

Ali, B., Green, K. M., Daughters, S. B., & Lejuez, C. (2017). Distress tolerance interacts with circumstances, motivation, and readiness to predict substance abuse treatment retention. *Addictive behaviors, 73*, 99-104.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Avila, C., & Parcet, M. A. (2001). Personality and inhibitory deficits in the stop-signal task: The mediating role of Gray's anxiety and impulsivity. *Personality and Individual Differences, 31*(6), 975-986.

Ball, S. A., Carroll, K. M., Canning-Ball, M., & Rounsaville, B. J. (2006). Reasons for dropout from drug abuse treatment: Symptoms, personality, and motivation. *Addictive behaviors, 31*(2), 320-330.

Ball, S. A., Martino, S., Nich, C., Frankforter, T. L., Van Horn, D., Crits-Christoph, P., . . . National Institute on Drug Abuse Clinical Trials, N. (2007). Site matters: multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *J Consult Clin Psychol, 75*(4), 556-567. doi:10.1037/0022-006X.75.4.556

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2015). Altering the cognitive-affective dysfunctions of psychopathic and externalizing offender subtypes with cognitive remediation. *Clinical Psychological Science, 3*(1), 45-57.

Baskin-Sommers, A. R., & Newman, J. P. (2013). Differentiating the Cognition-Emotion Interactions That Characterize Psychopathy versus Externalizing. In *Cognition and emotion* (pp. 501-520). New York: Guilford Press.

Beck, A. T., Emery G. (1985). Anxiety disorders and phobias. A cognitive perspective. *New York: Basic*.

Berg, C. J., Thomas, J. L., Guo, H., An, L. C., Okuyemi, K. S., Collins, T. C., & Ahluwalia, J. S. (2010). Predictors of smoking reduction among Blacks. *Nicotine & Tobacco Research, 12*(4), 423-431.

Berkman, E. T., Kahn, L. E., & Merchant, J. S. (2014). Training-induced changes in inhibitory control network activity. *Journal of Neuroscience, 34*(1), 149-157.

Bickel, W. K., Yi, R., Landes, R., Hill, P. F., & Baxter, C. (2011). Remembering the future: Working memory training decreases delay discounting among stimulant addicts. *Biol Psychiatry, 69*, 260-265.

Bjork, J. M., & Pardini, D. A. (2015). Who are those "risk-taking adolescents"? Individual differences in developmental neuroimaging research. *Dev Cogn Neurosci, 11*, 56-64. doi:10.1016/j.dcn.2014.07.008

Bornovalova, M. A., Gratz, K. L., Daughters, S. B., Hunt, E. D., & Lejuez, C. (2012). Initial RCT of a distress tolerance treatment for individuals with substance use disorders. *Drug and alcohol dependence, 122*(1-2), 70-76.

Brazil, I. A., van Dongen, J. D., Maes, J. H., Mars, R., & Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews, 91*, 259-277.

Brem, M. J., Florimbio, A. R., Elmquist, J., Shorey, R. C., & Stuart, G. L. (2018). Antisocial traits, distress tolerance, and alcohol problems as predictors of intimate partner violence in men arrested for domestic violence. *Psychology of violence, 8*(1), 132.

Brennan, G. M., Stuppy-Sullivan, A. M., Brazil, I. A., & Baskin-Sommers, A. R. (2017). Differentiating patterns of substance misuse by subtypes of antisocial traits in male offenders. *The Journal of Forensic Psychiatry & Psychology, 28*(3), 341-356.

Brorson, H. H., Arnevik, E. A., Rand-Hendriksen, K., & Duckert, F. (2013). Drop-out from addiction treatment: a systematic review of risk factors. *Clinical Psychology Review, 33*(8), 1010-1024.

Brown, R. A., Palm, K. M., Strong, D. R., Lejuez, C. W., Kahler, C. W., Zvolensky, M. J., . . . Gifford, E. V. (2008). Distress tolerance treatment for early-lapse smokers: Rationale, program description, and preliminary findings. *Behavior modification, 32*(3), 302-332.

Buckholtz, J. W. (2015). Social norms, self-control, and the value of antisocial behavior. *Current Opinion in Behavioral Sciences, 3*, 122-129. doi:10.1016/j.cobeha.2015.03.004

Buckholtz, J. W., Karmarkar, U., Ye, S., Brennan, G. M., & Baskin-Sommers, A. (2017). Blunted ambiguity aversion during cost-benefit decisions in antisocial individuals. *Scientific reports, 7*(1), 1-9.

Buckner, J. D., & Carroll, K. M. (2010). Effect of anxiety on treatment presentation and outcome: Results from the Marijuana Treatment Project. *Psychiatry Research, 178*(3), 493-500.

Byrd, A. L., Loeber, R., & Pardini, D. A. (2014). Antisocial behavior, psychopathic features and abnormalities in reward and punishment processing in youth. *Clinical Child and Family Psychology Review, 17*(2), 125-156.

Carroll, K. M., Ball, S. A., Martino, S., Nich, C., Babuscio, T. A., Nuro, K. F., . . . Rounsaville, B. J. (2008). Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. *Am J Psychiatry, 165*(7), 881-888. doi:10.1176/appi.ajp.2008.07111835

Carroll, K. M., Fenton, L. R., Ball, S. A., Nich, C., Frankforter, T. L., Shi, J., & Rounsaville, B. J. (2004). Efficacy of disulfiram and cognitive behavior therapy in cocaine-dependent outpatients: a randomized placebo-controlled trial. *Arch Gen Psychiatry, 61*(3), 264-272. doi:10.1001/archpsyc.61.3.264

Carroll, K. M., Nich, C., Petry, N. M., Eagan, D. A., Shi, J. M., & Ball, S. A. (2016). A randomized factorial trial of disulfiram and contingency management to enhance cognitive behavioral therapy for cocaine dependence. *Drug Alcohol Depend, 160*, 135-142. doi:10.1016/j.drugalcdep.2015.12.036

Chamberlain, S. R., Derbyshire, K. L., Leppink, E. W., & Grant, J. E. (2016). Neurocognitive deficits associated with Antisocial Personality Disorder in non-treatment-seeking young adults. *Journal of the American Academy of Psychiatry and the Law, 44*(2), 218-225.

Clark, D. A. (2013). Cognitive restructuring. *The Wiley handbook of cognitive behavioral therapy*, 1-22.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*, 386-396.

Compton, W. M., Cottle, L., Jacobs, J., Ben-Abdallah, A., & Spitznagel, E. (2003). The role of psychiatric disorders in predicting drug dependence treatment outcomes. *American Journal of Psychiatry, 160*(5), 890-895.

Craske, M. G. (2010). *Cognitive–behavioral therapy*: American Psychological Association.

Daughters, S. B., Lejuez, C., Bornovalova, M. A., Kahler, C. W., Strong, D. R., & Brown, R. A. (2005). Distress tolerance as a predictor of early treatment dropout in a residential substance abuse treatment facility. *Journal of Abnormal Psychology, 114*(4), 729-734.

Daughters, S. B., Lejuez, C. W., Kahler, C. W., Strong, D. R., & Brown, R. A. (2005). Psychological distress tolerance and duration of most recent abstinence attempt among residential treatment-seeking substance abusers. *Psychology of Addictive Behaviors, 19*(2), 208.

Daughters, S. B., Sargeant, M. N., Bornovalova, M. A., Gratz, K. L., & Lejuez, C. W. (2008). The relationship between distress tolerance and antisocial personality disorder among male inner-city treatment seeking substance users. *Journal of Personality Disorders, 22*(5), 509-524.

Davidson, K., Tyrer, P., Tata, P., Cooke, D., Gumley, A., Ford, I., . . . Robertson, H. (2009). Cognitive behaviour therapy for violent men with antisocial personality disorder in the community: an exploratory randomized controlled trial. *Psychological Medicine, 39*(4), 569.

Davidson, K. M., & Tyrer, P. (1996). Cognitive therapy for antisocial and borderline personality disorders: Single case study series. *British Journal of Clinical Psychology, 35*(3), 413-429.

De Brito, S. A., Viding, E., Kumari, V., Blackwood, N., & Hodgins, S. (2013). Cool and hot executive function impairments in violent offenders with antisocial personality disorder with and without psychopathy. *PLoS One, 8*(6), e65566.

De Leon, G., Melnick, G., Thomas, G., Kressel, D., & Wexler, H. K. (2000). Motivation for treatment in a prison-based therapeutic community. *The American journal of drug and alcohol abuse, 26*(1), 33-46.

Dentz, A., Guay, M.-C., Parent, V., & Romo, L. (2020). Working memory training for adults with ADHD. *Journal of attention disorders, 24*(6), 918-927.

Dolan, M. (2012). The neuropsychology of prefrontal function in antisocial personality disordered offenders with varying degrees of psychopathy. *Psychological Medicine, 42*(8), 1715-1725.

Dolan, M., & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy.

Dolan, M., & Park, I. (2002). The neuropsychology of antisocial personality disorder. *Psychological Medicine, 32*(3), 417.

Ducci, F., Enoch, M.-A., Funt, S., Virkkunen, M., Albaugh, B., & Goldman, D. (2007). Increased anxiety and other similarities in temperament of alcoholics with and without antisocial personality disorder across three diverse populations. *Alcohol, 41*(1), 3-12.

Duggan, C., Huband, N., Smailagic, N., Ferriter, M., & Adams, C. (2007). The use of psychological treatments for people with personality disorder: A systematic review of randomized controlled trials. *Personality and Mental Health, 1*(2), 95-125.

Estrada, S., Tillem, S., Stuppy-Sullivan, A., & Baskin-Sommers, A. (2019). Specifying the Connection Between Reward Processing and Antisocial Psychopathology Across Development. *The Oxford Handbook of Positive Emotion and Psychopathology*, 312.

Federal Bureau of Investigation. (2016). Crime in the U.S. 2015. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015

Felton, J. W., Collado, A., Ingram, K. M., Doran, K., & Yi, R. (2019). Improvement of working memory is a mechanism for reductions in delay discounting among mid-age individuals in an urban medically underserved area. *Annals of Behavioral Medicine, 53*(11), 988-998.

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). Structured clinical interview for DSM-5—Research version (SCID-5 for DSM-5, research version; SCID-5-RV). *Arlington, VA: American Psychiatric Association*.

Fonseca, A. C., & Yule, W. (1995). Personality and antisocial behavior in children and adolescents: an enquiry into Eysenck's and Gray's theories. *Journal of Abnormal Child Psychology, 23*(6), 767-781. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8609312

Gamito, P., Oliveira, J., Lopes, P., Brito, R., Morais, D., Silva, D., . . . Deus, A. (2014). Executive functioning in alcoholics following an mHealth cognitive stimulation program: randomized controlled trial. *Journal of medical Internet research, 16*(4), e102.

Gardiner, C., Tsukagoshi, S., Nur, U., & Tyrer, P. (2010). Associations of treatment resisting (Type R) and treatment seeking (Type S) personalities in medical students. *Personality and Mental Health, 4*(2), 59-63.

Gerst, K. R., Gunn, R. L., & Finn, P. R. (2017). Delay discounting of losses in alcohol use disorders and antisocial psychopathology: Effects of a working memory load. *Alcoholism: Clinical and Experimental Research, 41*(10), 1768-1774.

Gibbon, S., Khalifa, N. R., Cheung, N. H., Völlm, B. A., & McCarthy, L. (2020). Psychological interventions for antisocial personality disorder. *Cochrane Database of Systematic Reviews*(9).

Glassman, L. H., Martin, L. M., Bradley, L. E., Ibrahim, A., Goldstein, S. P., Forman, E. M., & Herbert, J. D. (2016). A brief report on the assessment of distress tolerance: Are we measuring the same construct? *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 34*(2), 87-99.

Glenn, A. L., Johnson, A. K., & Raine, A. (2013). Antisocial personality disorder: a current review. *Current Psychiatry Reports, 15*(12), 1-8. doi:10.1007/s11920-013-0427-7

Gratz, K. L., Bornovalova, M. A., Delany-Brumsey, A., Nick, B., & Lejuez, C. W. (2007). A laboratory-based study of the relationship between childhood abuse and experiential avoidance among inner-city substance users: The role of emotional nonacceptance. *Behavior therapy, 38*(3), 256-268.

Gunn, R. L., Gerst, K. R., Lake, A. J., & Finn, P. R. (2018). The effects of working memory load and attention refocusing on delay discounting rates in alcohol use disorder with comorbid antisocial personality disorder. *Alcohol, 66*, 9-14.

Gyurak, A., Gross, J. J., Chan, L., & Etkin, A. (2013). Cognitive-affective remediation training intervention in anxiety and depression. *Neuropsychopharmacology, 38*, S231.

Hajek, P., Belcher, M., & Stapleton, J. (1987). Breath-holding endurance as a predictor of success in smoking cessation. *Addictive behaviors, 12*(3), 285-288.

Hall, J. R., Bernat, E. M., & Patrick, C. J. (2007). Externalizing psychopathology and the error-related negativity. *Psychological Science, 18*(4), 326-333. doi:10.1111/j.1467-9280.2007.01899.x

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., & Scanlon, M. (2015). Enhancing cognitive abilities with comprehensive training: A large, online, randomized, active-controlled trial. *PLoS One, 10*(9), e0134467.

Harris, G. T., & Rice, M. E. (2006). Treatment of psychopathy. *Handbook of psychopathy*, 555-572.

Hatzitaskos, P., Soldatos, C. R., Kokkevi, A., & Stefanis, C. N. (1999). Substance abuse patterns and their association with psychopathology and type of hostility in male patients with borderline and antisocial personality disorder. *Comprehensive psychiatry, 40*(4), 278-282.

Hendershot, C. S., Wardell, J. D., Vandervoort, J., McPhee, M. D., Keough, M. T., & Quilty, L. C. (2018). Randomized trial of working memory training as an adjunct to inpatient substance use disorder treatment. *Psychology of Addictive Behaviors, 32*(8), 861.

Herrera, C., Chambon, C., Michel, B.-F., Paban, V., & Alescio-Lautier, B. (2012). Positive effects of computer-based cognitive training in adults with mild cognitive impairment. *Neuropsychologia, 50*(8), 1871-1881.

Herrera, P. M., Van Meerbeke, A. V., Speranza, M., Cabra, C. L., Bonilla, M., Canu, M., & Bekinschtein, T. A. (2019). Expectation of reward differentially modulates executive inhibition. *BMC psychology, 7*(1), 1-10.

Houben, K., Havermans, R. C., Nederkoorn, C., & Jansen, A. (2012). Beer à No-Go: Learning to stop responding to alcohol cues reduces alcohol intake via reduced affective associations rather than increased response inhibition. *Addiction, 107*(7), 1280-1287.

Houben, K., Nederkoorn, C., Wiers, R. W., & Jansen, A. (2011). Resisting temptation: decreasing alcohol-related affect and drinking behavior by training response inhibition. *Drug and alcohol dependence, 116*(1-3), 132-136.

Houben, K., Wiers, R. W., & Jansen, A. (2011). Getting a grip on drinking behavior: training working memory to reduce alcohol abuse. *Psychological Science, 22*(7), 968-975.

Jia, R., Ruderman, L., Gordon, C., Ehrlich, D., Horvath, M., Mirchandani, S., . . . Harpaz-Rotem, I. (2020). From value to saliency: neural computations of subjective value under uncertainty in combat veterans. *bioRxiv*.

Kendall, P. C., Finch, A. J., Jr., Auerbach, S. M., Hooke, J. F., & Mikulka, P. J. (1976). The State-Trait Anxiety Inventory: a systematic evaluation. *J Consult Clin Psychol, 44*(3), 406-412. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/932270

Kiluk, B. D., Devore, K. A., Buck, M. B., Nich, C., Frankforter, T. L., LaPaglia, D. M., . . . Carroll, K. M. (2016). Randomized Trial of Computerized Cognitive Behavioral Therapy for Alcohol Use Disorders: Efficacy as a Virtual Stand-Alone and Treatment Add-On Compared with Standard Outpatient Treatment. *Alcohol Clin Exp Res, 40*(9), 1991-2000. doi:10.1111/acer.13162

Kim, E. J., Bahk, Y.-C., Oh, H., Lee, W.-H., Lee, J.-S., & Choi, K.-H. (2018). Current status of cognitive remediation for psychiatric disorders: a review. *Frontiers in Psychiatry, 9*, 461.

Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental and Clinical psychopharmacology, 22*(3), 222.

Konova, A. B., Lopez-Guzman, S., Urmanche, A., Ross, S., Louie, K., Rotrosen, J., & Glimcher, P. W. (2020). Computational markers of risky decision-making for identification of

temporal windows of vulnerability to opioid use in a real-world clinical setting. *JAMA Psychiatry, 77*(4), 368-377.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139*(4), 665.

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2009). Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum.

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology, 116*(4), 645-666. doi:10.1037/0021-843x.116.4.645

Krueger, R. F., McGue, M., & Iacono, W. G. (2001). The higher-order structure of common DSM mental disorders: Internalization, externalization, and their connections to personality. *Personality and Individual Differences, 30*(7), 1245-1259.

Lappin, J. S., & Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *Journal of Experimental Psychology, 72*(6), 805.

Lejuez, C. W., Kahler, C. W., & Brown, R. A. (2003). A modified computer version of the Paced Auditory Serial Addition Task (PASAT) as a laboratory-based stressor. *Behav Ther, 26*, 290-293.

Leotti, L. A., & Wager, T. D. (2010). Motivational influences on response inhibition measures. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2), 430.

Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *J Neurophysiol, 103*(2), 1036-1047. doi:10.1152/jn.00853.2009

Lijffijt, M., Hu, K., & Swann, A. C. (2014). Stress modulates illness-course of substance use disorders: a translational review. *Frontiers in Psychiatry, 5*, 83.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological review, 91*(3), 295.

Manuel, A. L., Bernasconi, F., & Spierer, L. (2013). Plastic modifications within inhibitory control networks induced by practicing a stop-signal task: An electrical neuroimaging study. *Cortex, 49*(4), 1141-1147.

Martínez-Raga, J., Marshall, E. J., Keaney, F., Ball, D., & Strang, J. (2002). Unplanned versus planned discharges from in-patient alcohol detoxification: retrospective analysis of 470 first-episode admissions. *Alcohol and Alcoholism, 37*(3), 277-281.

Mazas, C. A., Finn, P. R., & Steinmetz, J. E. (2000). Decision-making biases, antisocial personality, and early-onset alcoholism. *Alcoholism: Clinical and Experimental Research, 24*(7), 1036-1040.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *Commons, ML.; Mazur, JE.; Nevin, JA*, 55-73.

McHugh, R. K., Daughters, S. B., Lejuez, C. W., Murray, H. W., Hearon, B. A., Gorka, S. M., & Otto, M. W. (2011). Shared variance among self-report and behavioral measures of distress intolerance. *Cognitive therapy and research, 35*(3), 266-275.

Medalia, A., & Bowie, C. R. (2016). *Cognitive remediation to improve functional outcomes*: Oxford University Press.

Mellis, A. M., Woodford, A. E., Stein, J. S., & Bickel, W. K. (2017). A second type of magnitude effect: Reinforcer magnitude differentiates delay discounting between substance users and controls. *Journal of the Experimental Analysis of Behavior, 107*(1), 151-160.

Messina, N., Farabee, D., & Rawson, R. (2003). Treatment responsivity of cocaine-dependent patients with antisocial personality disorder to cognitive-behavioral and contingency management interventions. *Journal of Consulting and Clinical Psychology, 71*(2), 320.

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., . . . DeRubeis, R. J. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature reviews Drug discovery, 11*(2), 141-168.

Miller, T. R., Cohen, M. A., Swedler, D. I., Ali, B., & Hendrie, D. V. (2020). Incidence and Costs of Personal and Property Crimes in the USA, 2017. *Journal of Benefit-Cost Analysis,* 1-31.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.

Moody, L., Franck, C., & Bickel, W. K. (2016). Comorbid depression, antisocial personality, and substance dependence: Relationship with delay discounting. *Drug and alcohol dependence, 160*, 190-196.

National Collaborating Centre for Mental Health. (2010). Interventions for people with antisocial personality disorder and associated symptoms and behaviors. In *Antisocial Personality Disorder: Treatment, Management and Prevention*: British Psychological Society.

National Institute on Drug Abuse. (2017, April 21, 2017). Trends & Statistics. Retrieved from https://www.drugabuse.gov/related-topics/trends-statistics

Ogilvie, J. M., Stewart, A. L., Chan, R. C. K., & Shum, D. H. K. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology, 49*(4), 44.

Onken, L. S., Carroll, K. M., Shoham, V., Cuthbert, B. N., & Riddle, M. (2014). Reenvisioning clinical science unifying the discipline to improve the public health. *Clinical Psychological Science, 2*(1), 22-34.

Patrick, C. J., Durbin, C. E., & Moser, J. S. (2012). Reconceptualizing antisocial deviance in neurobehavioral terms. *Development and Psychopathology, 24*(3), 1047-1071. doi:10.1017/s0954579412000533

Petry, N. M. (2002). Discounting of delayed rewards in substance abusers: relationship to antisocial personality disorder. *Psychopharmacology, 162*(4), 425-432.

Pew Research Center. (2018). Mobile fact sheet. *Pew Research Center*.

Prehn, K., Schulze, L., Rossmann, S., Berger, C., Vohs, K., Fleischer, M., . . . Herpertz, S. C. (2013). Effects of emotional stimuli on working memory processes in male criminal offenders with borderline and antisocial personality disorder. *The World Journal of Biological Psychiatry, 14*(1), 71-78.

Purves, D., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., Woldorff, M. G., & Brannon, E. M. (2008). *Cognitive Neuroscience*: Sunderland: Sinauer Associates, Inc.

Raine, A. (2018). Antisocial personality as a neurodevelopmental disorder. *Annual Review of Clinical Psychology, 14*, 259-289.

Raine, A., & Scerbo, A. (1991). Biological theories of violence. In *Neuropsychology of aggression* (pp. 1-25): Springer.

Rass, O., Schacht, R. L., Buckheit, K., Johnson, M. W., Strain, E. C., & Mintzer, M. Z. (2015). A randomized controlled trial of the effects of working memory training in methadone maintenance patients. *Drug and alcohol dependence, 156*, 38-46.

Robinson, S. M., Sobell, L. C., Sobell, M. B., & Leo, G. I. (2012). Reliability of the Timeline Followback for Cocaine, Cannabis, and Cigarette Use. *Psychol Addict Behav*. doi:10.1037/a0030992

Roos, C. R., Kiluk, B. D., McHugh, R. K., & Carroll, K. M. (2020). Evaluating a longitudinal mediation model of perceived stress, depressive symptoms, and substance use treatment outcomes. *Psychology of Addictive Behaviors*.

Rowe, D. C. (1997). Are parents to blame? A look at the antisocial personalities. *Psychological Inquiry, 8*(3), 251-260.

Rubio, G., Jimenez, M., Rodriguez-Jimenez, R., Martinez, I., Iribarren, M. M., Jimenez-Arriero, M. A., . . . Avila, C. (2007). Varieties of impulsivity in males with alcohol dependence: the role of Cluster-B personality disorder. *Alcoholism: Clinical and Experimental Research, 31*(11), 1826-1832. doi:10.1111/j.1530-0277.2007.00506.x

Rubio, G., Jiménez, M., Rodríguez-Jiménez, R., Martínez, I., Iribarren, M. M., Jiménez-Arriero, M. A., . . . Avila, C. (2007). Varieties of impulsivity in males with alcohol dependence: the role of cluster-B personality disorder. *Alcoholism: Clinical and Experimental Research, 31*(11), 1826-1832.

Rupp, C. I., Kemmler, G., Kurz, M., Hinterhuber, H., & Wolfgang Fleischhacker, W. (2012). Cognitive remediation therapy during treatment for alcohol dependence. *Journal of Studies on Alcohol and Drugs, 73*(4), 625-634.

Salekin, R. T. (2002). Psychopathy and therapeutic pessimism: Clinical lore or clinical reality? *Clinical Psychology Review, 22*(1), 79-112.

Sargeant, M. N., Daughters, S. B., Curtin, J. J., Schuster, R. M., & Lejuez, C. (2011). Unique roles of antisocial personality disorder and psychopathic traits in distress tolerance. *Journal of Abnormal Psychology*, *120*(4), 987.

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European archives of psychiatry and clinical neuroscience, 264*(1), 61-69.

Schuckit, M. A., Tapert, S., Matthews, S. C., Paulus, M. P., Tolentino, N. J., Smith, T. L., . . . Simmons, A. (2012). fMRI differences between subjects with low and high responses to alcohol during a stop signal task. *Alcoholism: Clinical and Experimental Research, 36*(1), 130-140.

Spunt, R. P., Lieberman, M. D., Cohen, J. R., & Eisenberger, N. I. (2012). The phenomenology of error processing: the dorsal ACC response to stop-signal errors tracks reports of negative affect. *Journal of Cognitive Neuroscience, 24*(8), 1753-1765.

Stark, M. J. (1992). Dropping out of substance abuse treatment: A clinically oriented review. *Clinical Psychology Review, 12*(1), 93-116.

Stevens, M. C., Kaplan, R. F., & Hesselbrock, V. M. (2003). Executive–cognitive functioning in the development of antisocial personality disorder. *Addictive behaviors, 28*(2), 285-300.

Strong, D. R., Lejuez, C. W., Daughters, S., Marinello, M., Kahler, C. W., & Brown, R. A. (2003). The computerized mirror tracing task, version 1. *Unpublished manual*.

Sutterlin, S., Schroijen, M., Constantinou, E., Smets, E., Van den Bergh, O., & Van Diest, I. (2013). Breath holding duration as a measure of distress tolerance: examining its relation to measures of executive control. *Front Psychol, 4*, 483. doi:10.3389/fpsyg.2013.00483

Swann, A. C., Lijffijt, M., Lane, S. D., Steinberg, J. L., & Moeller, F. G. (2009). Trait impulsivity and response inhibition in antisocial personality disorder. *Journal of Psychiatric Research, 43*(12), 1057-1063.

Thevos, A. K., Thomas, S. E., & Randall, C. L. (1999). Baseline differences in social support among treatment-seeking alcoholics with and without social phobia. *Substance Abuse, 20*(2), 107-118.

Thomas, S. E., Thevos, A. K., & Randall, C. L. (1999). Alcoholics with and without social phobia: a comparison of substance use and psychiatric variables. *Journal of Studies on Alcohol, 60*(4), 472-479.

Thylstrup, B., & Hesse, M. (2016). Impulsive lifestyle counseling to prevent dropout from treatment for substance use disorders in people with antisocial personality disorder: A randomized study. *Addictive behaviors, 57*, 48-54.

Tomlinson, D. C., Tegge, A. N., Athamneh, L. N., & Bickel, W. K. (2020). The phenotype of recovery IV: Delay discounting predicts perceived stress and a chance locus of control in individuals in recovery from substance use disorders. *Addictive Behaviors Reports, 12*, 100320.

Turner, T. H., LaRowe, S., Horner, M. D., Herron, J., & Malcolm, R. (2009). Measures of cognitive functioning as predictors of treatment outcome for cocaine dependence. *Journal of substance abuse treatment, 37*(4), 328-334.

Tyrer, P., Mitchard, S., Methuen, C., & Ranger, M. (2003). Treatment rejecting and treatment seeking personality disorders: Type R and Type S. *Journal of Personality Disorders, 17*(3), 263-268.

Verbruggen, F., Chambers, C. D., & Logan, G. D. (2013). Fictitious inhibitory differences: how skewness and slowing distort the estimation of stopping latencies. *Psychological Science, 24*(3), 352-362.

Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-IT: Windows executable software for the stop-signal paradigm. *Behavior research methods, 40*(2), 479-483.

Verdejo-Garcia, A. (2016). Cognitive training for substance use disorders: Neuroscientific mechanisms. *Neuroscience & Biobehavioral Reviews, 68*, 270-281.

Vince, M. A. (1948). The intermittency of control movements and the psychological refractory period. *British Journal of Psychology, 38*(3), 149-157.

Warren, F., Preedy-Fayers, K., McGauley, G., Pickering, A., Norton, K., Geddes, J., & Dolan, B. (2003). Review of treatments for severe personality disorder. *Home Office online report, 30*(03).

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19*(1), 87-95.

Wei, L. J., & Lachin, J. M. (1988). Properties of the urn randomization in clinical trials. *Control Clin Trials, 9*(4), 345-364. Retrieved from

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3203525

White, S. F., Clanton, R., Brislin, S. J., Meffert, H., Hwang, S., Sinclair, S., & Blair, R. J. R. (2014). Reward: Empirical contribution: Temporal discounting and conduct disorder in adolescents. *Journal of Personality Disorders*, *28*(1), 5-18.

Wölwer, W., Burtscheidt, W., Redner, C., Schwarz, R., & Gaebel, W. (2001). Out-patient behaviour therapy in alcoholism: impact of personality disorders and cognitive impairments. *Acta Psychiatrica Scandinavica*, *103*(1), 30-37.

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. *American Journal of Psychiatry*, *168*(5), 472-485.

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised Manual*. Los Angeles, CA: Western Psychological Services.

Zeier, J. D., Baskin-Sommers, A. R., Hiatt Racer, K. D., & Newman, J. P. (2012). Cognitive control deficits associated with antisocial personality disorder and psychopathy. *Personality Disorders: Theory, Research, and Treatment*, *3*(3), 283.

Zvolensky, M. J., Feldner, M. T., Eifert, G. H., & Brown, R. A. (2001). Affective style among smokers: Understanding anxiety sensitivity, emotional reactivity, and distress tolerance using biological challenge. *Addictive behaviors*, *26*(6), 901-915.

# Chapter 5: General Discussion

The research presented in this dissertation provides specificity to accounts of cognitive-affective dysfunctions impacting antisocial individuals and illustrates how designing treatments with these cognitive-affective dysfunctions in mind can propel behavior change in a population that is often deemed to be treatment-resistant. Findings from Studies 1 and 2 highlighted how executive functioning, decision-making, and reward sensitivity are not globally deficient among antisocial individuals. Instead, multiple cognitive processes (perception, inhibition under high load, executive functioning more generally, effort-based decision-making) were disrupted in certain affective circumstances (e.g., when high value rewards were at stake, when individuals were aware of high value rewards, when individuals were experiencing high levels of negative affect), ultimately influencing how antisocial individuals processed information and responded behaviorally. Findings from Study 3 indicated that specific cognitive-affective dysfunctions identified in previous studies of antisocial individuals, including studies from this dissertation, could be addressed directly through targeted cognitive remediation training, leading to improvements in both cognition and behavior. Specifically, antisocial individuals enrolled in outpatient substance use treatment were capable of enhancing their ability to inhibit responses in both neutral and rewarding contexts, and their ability to update working memory in distressing contexts. Further, these advancements led to improvements in working memory capability, substance use frequency, and therapy session attendance. Overall, results demonstrated that refining conceptualizations of cognitive-affective dysfunctions in antisocial individuals provides a more accurate account of their engagement in maladaptive behavior patterns and, ultimately, can facilitate a reversal of these behavior patterns by informing targeted interventions.

Results from this dissertation have important implications when considered alongside traditional accounts of antisocial behavior and attitudes toward remediating it. Traditional

models of antisocial behavior most often focus broadly on diminished executive functioning and aberrant decision-making. Further, several theories and experimental studies of antisocial behavior describe how abnormal reward sensitivity may impact executive functions and decision-making. The present findings highlight anomalies in the perceptual encoding of information, the integration of information during effort-based decision-making, and the impact of distress in disrupting cognition and behavior in antisocial individuals. Abnormal perception in the presence of rewards may contribute to executive functioning and decision-making abnormalities documented in prior studies, since accurately perceiving one's environment is a prerequisite for detecting cues signaling that inhibition is necessary, or accurately evaluating information about choice options ahead of action selection. Additionally, while reward is influential in antisocial behavior (Estrada, Tillem, Stuppy-Sullivan, & Baskin-Sommers, 2019), so too is negative affect, which impacts whether antisocial individuals can integrate information about rewards when choosing how to invest their efforts. An approach to understanding antisocial behavior that incorporates additional domains of cognition and affect sheds light on the extent of dysfunction in antisocial individuals and provides novel targets for interventions.

Although findings from this dissertation suggest more widespread dysfunction in antisocial individuals than traditional accounts would indicate, they also suggest instances of intact cognitive-affective capacities that can be leveraged for treatment. For example, Study 1 indicated that antisocial individuals do not show inhibition problems when rewards at stake and working memory demands are relatively low. Additionally, Studies 1 and 2 indicated that antisocial individuals can integrate information about reward values and probabilities during risk-based decision-making, and even during effort-based decision-making if they are experiencing lower levels of negative affect. Study 3 suggested that when given opportunities to

practice specific skills in progressively more challenging environments, antisocial individuals can build toward improving areas of relative weakness. Specifically, antisocial individuals can gain practice with inhibition under varying levels of reward (e.g., practicing in both neutral and rewarded contexts) and difficulty (e.g., in response to varying Stop Signal Delays), as well as gain exposure integrating information while under progressively higher levels of distress (e.g., with decreased latency between stimuli during the PASAT-C). Identifying areas within executive functioning, decision-making, and reward processing that are intact in antisocial individuals provides opportunities to bolster cognitive-affective functioning and behavioral change by leveraging these capabilities to address specific dysfunctions in antisocial individuals.

Findings from this dissertation represent a step toward describing cognitive-affective dysfunctions and capacities in antisocial individuals more thoroughly than previous research and improving treatments based on this knowledge. However, further programmatic research related to cognitive-affective functioning and antisocial individuals is needed. Future efforts to understand and treat antisocial individuals will benefit from exploring additional subfactors of cognitive (e.g., attention, social decision-making) and affective (e.g., threat, loss, frustrative non-reward, boredom) processes, as well as how such processes relate to one another. Basic science underscores important distinctions within cognitive and affective domains (e.g., the multiple well-defined and empirically supported factors subsumed by "impulsivity;" Strickland & Johnson, 2021) and important moderators of cognitive-affective functioning in neurotypical individuals. By extension, the translation of basic science to clinical science is needed to examine potential roles for specific cognitive-affective factors, subfactors, and moderators in the behavior of antisocial individuals more fully. For example, future studies may examine whether abnormal perception in highly rewarding contexts underlies abnormal social perception among

antisocial individuals (e.g., contributes to abnormal processing of angry or fearful faces), and whether such disruption is particularly prominent among antisocial individuals who engage in violent crimes. Such findings may suggest a role for mindfulness-based interventions emphasizing visual processing with progressively more complex and affectively charged stimuli in treating more violent antisocial individuals, potentially reducing the public health burden associated interpersonal violence (Corso, Mercy, Simon, Finkelstein, & Miller, 2007). With a larger collection of findings documenting specific dysfunctional versus intact processes, researchers focusing on antisocial individuals can formulate a comprehensive conceptualization of the etiopathogenesis of antisocial behavior and treatments that leverage and target processes that can positively impact behavior.

Future work may hone conceptualizations of antisocial individuals and increase treatment potency further by not only attending more to their cognitive-affective dysfunctions and capabilities, but also by incorporating understandings of additional factors conferring risk for continued antisocial behavior. Environmental (e.g., poverty; Chung, 2004; Machell, Disabato, & Kashdan, 2016), interpersonal (e.g., conflicts with treatment providers; Ball, Carroll, Canning-Ball, & Rounsaville, 2006), and co-morbid psychiatric (e.g., Post-traumatic stress disorder; Ardino, 2012) factors are influential in the onset and maintenance of antisocial behavior. Additionally, environmental factors, interpersonal experiences, and psychiatric co-morbidities are likely to interact with cognitive-affective functions, impacting both day-to-day functioning and treatment engagement (see Ford, 2015; Levi, Laslo-Roth, & Rosenstreich, 2018; for reviews related to trauma and poverty, respectively). For example, future studies might examine whether cognitive load and stress created by financial constraints (Mani, Mullainathan, Shafir, & Zhao, 2013; Santiago, Wadsworth, & Stump, 2011; Shafir, 2017; Shah, Zhao, Mullainathan, & Shafir,

2018) uniquely impacts inhibition under reward or threat among antisocial individuals who face poverty, and whether therapeutic efforts to address financial constraints (e.g., providing vocational services, covering transportation costs for attending therapy) can buffer against any inhibition failures uniquely influencing behavior or treatment engagement for antisocial individuals impacted by poverty. On the whole, antisocial behavior may be "difficult to treat" (Paris, Chenard-Poirier, & Biskin, 2013, p. 323) because it is difficult to understand, reflecting a complex intersection of cognitive-affective, environmental, and psychological factors. Nonetheless, additional insight into the multi-factorial causes of antisocial behavior can help replace unsuccessful "one-size fits all" approaches (Snyder, Reid, & Patterson, 2003, p. 40) by informing personalized treatment programs to fit the diverse needs of antisocial individuals (Brazil, van Dongen, Maes, Mars, & Baskin-Sommers, 2018).

In conclusion, refining mechanistic research and its connection to applied research has the power to revolutionize behavior change, yielding the most potent interventions possible for even the most chronic and severe antisocial individuals (Onken, Carroll, Shoham, Cuthbert, & Riddle, 2014). Continued research is needed to develop interventions that target specific mechanisms, evaluate changes in these specific mechanisms, and examine the extent to which observed changes generalize to promote changes in cognitive-affective and behavioral functioning. This systematic approach promotes the personalization of behavior change by integrating work across theoretical and methodological domains. Although conducting more rigorous mechanistic work and mechanism-informed applied work may be challenging, such efforts are worthwhile in their potential to alleviate the enormous suffering caused by antisocial behavior.

## References: General Discussion

Ardino, V. (2012). Offending behaviour: The role of trauma and PTSD. In: Taylor & Francis.

Ball, S. A., Carroll, K. M., Canning-Ball, M., & Rounsaville, B. J. (2006). Reasons for dropout from drug abuse treatment: Symptoms, personality, and motivation. *Addictive behaviors, 31*(2), 320-330.

Brazil, I. A., van Dongen, J. D., Maes, J. H., Mars, R., & Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews, 91*, 259-277.

Chung, I.-J. (2004). A conceptual framework for understanding the relationship between poverty and antisocial behavior: Focusing on psychosocial mediating mechanisms. *Journal of Primary Prevention, 24*(3), 375-400.

Corso, P. S., Mercy, J. A., Simon, T. R., Finkelstein, E. A., & Miller, T. R. (2007). Medical costs and productivity losses due to interpersonal and self-directed violence in the United States. *American Journal of Preventive Medicine, 32*(6), 474-482. doi:10.1016/j.amepre.2007.02.010

Estrada, S., Tillem, S., Stuppy-Sullivan, A., & Baskin-Sommers, A. (2019). Specifying the Connection Between Reward Processing and Antisocial Psychopathology Across Development. *The Oxford Handbook of Positive Emotion and Psychopathology*, 312.

Ford, J. (2015). An affective cognitive neuroscience-based approach to PTSD psychotherapy: The TARGET model. *Journal of Cognitive Psychotherapy, 29*(1), 68-91.

Levi, U., Laslo-Roth, R., & Rosenstreich, E. (2018). Socioeconomic Status and Psychotherapy: A Cognitive-Affective View. *J Psychiatry Behav Health Forecast. 2018; 1 (2), 1008*.

Machell, K. A., Disabato, D. J., & Kashdan, T. B. (2016). Buffering the negative impact of poverty on youth: The power of purpose in life. *Social Indicators Research, 126*(2), 845-861.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science, 341*(6149), 976-980.

Onken, L. S., Carroll, K. M., Shoham, V., Cuthbert, B. N., & Riddle, M. (2014). Reenvisioning clinical science unifying the discipline to improve the public health. *Clinical Psychological Science, 2*(1), 22-34.

Paris, J., Chenard-Poirier, M.-P., & Biskin, R. (2013). Antisocial and borderline personality disorders revisited. *Comprehensive psychiatry, 54*(4), 321-325.

Santiago, C. D., Wadsworth, M. E., & Stump, J. (2011). Socioeconomic status, neighborhood disadvantage, and poverty-related stress: Prospective effects on psychological syndromes among diverse low-income families. *Journal of Economic Psychology, 32*(2), 218-230.

Shafir, E. (2017). Decisions in poverty contexts. *Current opinion in psychology, 18*, 131-136.

Shah, A. K., Zhao, J., Mullainathan, S., & Shafir, E. (2018). Money in the mental lives of the poor. *Social Cognition, 36*(1), 4-19.

Snyder, J., Reid, J., & Patterson, G. (2003). A social learning model of child and adolescent antisocial behavior. *Causes of conduct disorder and juvenile delinquency*, 27-48.

Strickland, J. C., & Johnson, M. W. (2021). Rejecting impulsivity as a psychological construct: A theoretical, empirical, and sociocultural argument. *Psychological review, 128*(2), 336.

# Appendix A: Study 1 Supplemental Material

## Supplemental Methods

### Participants

Participants were 116 male inmates from a maximum-security correctional institution in Connecticut (see Table 1 for sample characteristics). A prescreen of institutional files and assessment materials were used to exclude individuals who: were not between the ages of 18 and 75, scored below 70 on a brief measure of IQ (Shipley Institute of Living Scale; Zachary, 1986) , performed below the fourth-grade level on a standardized measure of reading (Wide Range Achievement Test-III; Wilkinson, 1993), had diagnoses of schizophrenia, bipolar disorder, psychosis not otherwise specified, were currently taking psychotropic medication, or had a history of medical problems (e.g., uncorrectable auditory or visual deficits, head injury with loss of consciousness greater than 30 minutes, seizures) that could impact their comprehension of the study materials. Participants completed a semi-structured diagnostic interview on one visit and the three laboratory tasks on a second visit. Across the three tasks, participants were instructed to attend to reward cues, which might be hard to see at times, and to try to earn as many reward points as possible in order to be added to a "leader board" that was on display to all study participants in the testing room (*Note*: Connecticut Department of Correction does not allow researchers to pay inmates). All participants were provided written informed consent according to the procedures set forth by the Yale University Human Investigation Committee.

### Measures

**Antisocial Personality Disorder (APD).** Participants were assessed for APD during a semi-structured diagnostic interview. The interview evaluated the age and frequency of engagement in behaviors outlined in the Diagnostic Statistical Manual-5 (DSM-5; American

Psychiatric Association, 2013). A diagnosis of APD was given if there was evidence of conduct disorder (CD) prior to age 15 (three or more symptoms) and sufficient adult antisocial symptoms (three or more). Inter-rater reliability for 32% of the sample was .989 (Cohen's kappa).

**Masked Reward Cues.** Before each trial in the three tasks, the point value at stake for the trial was displayed using a modified reward masking paradigm (see Figure 1; Bijleveld et al., 2009). Point values were low (1 point) or high (10 points), noted by blocked digits (01 and 10, respectively). These reward cues were displayed either consciously (i.e., for a duration that is consciously perceivable, 300ms) or unconsciously (i.e., 30ms). Immediately before and after the presentation of the reward cues, masks consisting of overlapping block 0's and 1's were presented (100ms and 235ms for conscious and unconscious reward cues, respectively; total time for each masked cue was held at 500ms). Participants were told to look for reward cues before each trial to find out how many points were at stake. Participants were reminded that these point values might be difficult to see at times. Masked reward cues always appeared between fixation crosses on a black background, with 700-1200ms of total fixation, for a total cue and fixation period from 1200-1700ms (1450ms average).

**Visual search task.** For the perception task, a modified version of a visual search task was used (Kristjánsson et al., 2010) (see Figure 2A). During the task, participants viewed a series of displays with three colored diamonds. Participants were instructed to search for the oddly colored diamond, either a red target among two green distractors or vice versa. Participants indicated (by button press) whether the oddly colored diamond had a notch missing at the top or the bottom of the shape.

Every trial started with a masked reward cue described above. Following the masked reward cue, a display containing three diamond shapes was presented for 1000ms. Participants were allowed to provide a response for the duration (1000ms) the diamonds were on the screen. Participants received feedback (correct/incorrect and the number of points earned) after each trial (1000ms). After feedback, the next trial continued immediately, starting with a masked reward cue.

A total of 400 trials were presented in blocks of 40 trials (10 blocks total). Across all trials, diamond position, notch position, and pop-out color were counterbalanced (approximately 33 per subtype of trial). Reward cues (magnitude and consciousness) also were counterbalanced (100 trials per magnitude by consciousness condition).

Since performance for this task may include changes in speed or accuracy, an inverse efficiency score (IES; mean response time for correct responses divided by percentage of correct responses) was calculated for each participant. All participants obtained accuracy greater than 80% and were included in the analyses, resulting in a sample of 116 participants for the visual search task.

***n*-back task.** For the executive functioning task, we used a modified version of the *n*-back task (see Figure 2B; Baskin-Sommers et al., 2014, Pochon et al., 2002). During the task, participants viewed a series of letters. Participants were instructed to monitor the letters and respond with a button press if the preceding letter in the *n*-back position was different from the current letter (e.g., a mismatch trial). Participants were instructed to withhold their responses when the preceding letter matched the current stimulus (e.g., a match trial). The majority of trials were mismatch trials (80%), whereas match trials were infrequent (occurring 20% of the time). The task also included a manipulation of working memory load. In the low load (1-back)

condition, participants were instructed to determine whether the currently presented letter matched the immediately preceding letter in the sequence. In the high load (2-back) condition, participants were required to monitor and maintain the stimulus information in working memory in order to determine whether the letter stimulus 2 positions earlier matched the current letter.

Every trial started with a masked reward cue described above. Following the masked reward cue, participants viewed a series of 12 letters presented for 500ms each with an inter-letter-interval of 2000ms. For each letter, participants were allowed to provide a response anytime during the duration of the letter presentation or during the inter-letter-interval (2500ms response window). Participants received feedback (correct/incorrect and the number of points earned) for 2000ms after each trial (i.e., string of 12-letters). Following feedback, a blank screen was presented for a 2000ms intertrial interval.

There were a total of 40 trials (i.e., 12-letter runs). Trials were blocked into five runs of the same load level, with reward cue magnitude (low vs. high) and consciousness (unconscious/30ms vs. conscious/300ms) varying between runs. Therefore, a total of ten 12-letter trials were presented per reward by consciousness condition, split between low and high load.

For each participant, accuracy on the task was calculated. All 116 participants completed the *n*-back task, however, seven participants performed below 40% accuracy on low load match trials or below 20% accuracy on high load match trials. Data for these seven participants were excluded from the main analyses, resulting in a sample of 109 participants for the *n*-back task.

**Gambling task**. To assess probabilistic decision-making, a gambling task was used to examine risk taking behavior (modified gain conditions from Voon et al., 2006; see Figure 2C). During the task, participants viewed a series of two circles (i.e., gamble options). Participants were instructed to make a choice between one of two gamble options: a 'sure' and a 'risky' option. Participants were to press the right button for the option on the right of the screen and left button for the option on the left of the screen.

Every trial started with a masked reward cue described above. Following each masked reward cue, participants viewed the two gamble options for 4500ms. Participants were able to make a choice at any time during the gamble display. One 'sure' option always provided a small but certain number of points (5-113 points, depicted as a circle filled in one color with one number of points in the center), while the other 'risky' option always provided a chance (P = .35, .40, or .45) of winning some amount of points (range 15-250 points), and a chance of winning zero points (depicted as a circle with two segments proportionate to win and no-win probabilities, with the associated numbers of points at the center of each segment). Participants were informed that for 10-point trials, the reward values presented in the decision-making task would be multiplied by ten (e.g., a choice between a certain 25 points and a 40% chance of gaining 55 points would really be a choice between a certain 250 points and a 40% chance of gaining 550 points). 'Sure' and 'risky' options were matched for expected value (i.e., each of the two options was equally rewarding when considering both value and probability). Participants received feedback (i.e., points earned) for 1000ms. After feedback, the next trial continued immediately, starting with a masked reward cue.

A total of 160 trials were presented. There were 36 trials for each reward by consciousness condition. Additionally, there were 16 catch trials. For catch trials, the 'risky'

option was always worse than the 'sure' option (e.g. 'sure' choice of 62 points versus 'risky' choice with a 20% chance to win 25 points) to ensure subjects were paying attention to and understood the task.

For each participant, the percentage of 'risky' choices was calculated. All 116 participants chose the higher expected value option in over 60% of the catch trials; therefore, data for all 116 participants were analyzed for the decision-making task.

## Supplemental Results

### Performance across tasks

APD was related to worse performance for high compared to low rewards during the visual search perception task and the high-load inhibition trials of the *n*-back task. To examine whether it was the same individuals who showed poorer performance in response to high rewards across these tasks, and if performance across these tasks was related to APD, we examined behavior for the 109 participants who completed both the visual search and *n*-back tasks. First, we computed difference scores reflecting the extent to which individuals performed poorly for high vs. low rewards in each task (i.e., higher visual search inverse efficiency and lower *n*-back inhibition accuracy during high load). Second, we created four bins for each task difference score based on +/- 1 SD and summed these binned groups, such that higher bin sums reflected worse performance for high vs. low rewards across both tasks. Third, from these sums we divided participants into three groups, with individuals in the first group having the lowest bin sums (least impaired performance for high vs. low rewards), individuals in the second group having moderate sums (reflecting moderate declines in performance for high vs. low rewards), and individuals in the third group having the highest bin sums (worst performance for high vs. low

rewards across both tasks). Finally, we created dummy-coded variables to represent membership in the three groups, and conducted a binomial logistic regression.

The binomial logistic regression model provided good fit for the data [$N = 109$; $R^2 = .06$ (Cox & Snell); $R^2 = .07$ (Nagelkerke); Model $\chi^2(2) = 6.249$; $p = .044$]. Individuals in the third group, who were characterized by poor performance for high rewards across both tasks, were 1.46 times more likely to have APD than individuals in the first group, who demonstrated the least impaired performance for high vs. low rewards in the two laboratory tasks, OR = 1.46, $p = .018$, 95% CI [1.07, 1.99]. Individuals in the second group, characterized by moderate performance for high rewards across both tasks, were no more likely to have APD compared to individuals in the first group, OR = 1.47, $p = .104$, 95% CI [0.92, 2.35]. These results suggest that individuals with APD performed poorly in response to high rewards across both tasks, and that individuals with this behavioral pattern were more likely to have APD than individuals who showed less impairment in response to high rewards in these tasks.

**Supplemental References: Study 1**

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)* (5th, text revision ed.). Washington, D. C.: American Psychiatric Pub.

Baskin-Sommers, A. R., Krusemark, E. A., Curtin, J. J., Lee, C., Vujnovich, A., & Newman, J. P. (2014). The impact of cognitive control, incentives, and working memory load on the P3 responses of externalizing prisoners. *Biological Psychology, 96*, 86-93. doi:10.1016/j.biopsycho.2013.12.005

Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological Science, 20*(11), 1313-1315.

Kristjánsson, Á., Sigurjónsdóttir, Ó., & Driver, J. (2010). Fortune and reversals of fortune in visual search: Reward contingencies for pop-out targets affect search efficiency and target repetition effects. *Attention, Perception, & Psychophysics, 72*(5), 1229-1236.

Pochon, J., Levy, R., Fossati, P., Lehericy, S., Poline, J., Pillon, B., . . . Dubois, B. (2002). The neural system that bridges reward and cognition in humans: an fMRI study. *Proceedings of the National Academy of Sciences, 99*(8), 5669-5674.

Voon, V., Hassan, K., Zurowski, M., Duff-Canning, S., De Souza, M., Fox, S., . . . Miyasaki, J. (2006). Prospective prevalence of pathologic gambling and medication association in Parkinson disease. *Neurology, 66*(11), 1750-1752.

Wilkinson, G. S. (1993). *The Wide Range Achievement Test: Manual* (3rd ed.). Wilmington, DE: Wide Range, Inc.

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised Manual*. Western Psychological Services, Los Angeles, CA.

# Appendix B: Study 2 Supplemental Material

## Supplemental Method

### Measures

**Structured Clinical Interview for DSM-5 Disorders** (SCID-5; First, Williams, Karg, & Spitzer, 2015). Abnormal willingness to exert effort to obtain rewards is considered a transdiagnostic marker of psychiatric risk and illness (Barch, Treadway, & Schoen, 2014; Treadway, Bossaller, Shelton, & Zald, 2012; Treadway, Buckholtz, Schwartzman, Lambert, & Zald, 2009). In fact, the EEfRT has been used extensively in populations diagnosed with MDD. Given the substantial comorbidity between APD and MDD (Khan, Jacobson, Gardner, Prescott, & Kendler, 2005), we examined whether MDD symptoms would impact the relationship between EV and antisociality. The SCID was used to determine current symptoms of Major Depressive Disorder (MDD). Current MDD symptoms were summed. In Supplemental Table S1 we list MDD and other diagnoses assessed using the SCID-5 that characterized the present sample.

**Trail Making Test-B** (Allen & Haderlie, 2010). Successfully integrating multiple sources of information into an expected value signal to be used to guide effort requires an ability to understand and work with numbers. The Trail Making Test-B provided a measure of numeracy and helped to rule out that diminished EV use was related to poorer numeracy rather than level of APD symptoms. The Trail Making Test-B requires participants to connect 25 numbers and letters, in sequence, alternating between numbers and letters (e.g. 1, A, 2, B, 3, C). Test administraters monitor and time participants. Shorter task completion times indicate higher numeracy, along with better general intellectual functioning, and were normalized using natural-log transformations.

**Digits Backward** (Wechsler, 1945)**.** Maintaining a mental representation of reward value and probability on the EEfRT may rely on working memory processes (Treadway et al., 2015). To rule out whether integration deficits associated with APD symptoms truly reflected working memory deficits, which are common in individuals with APD (De Brito, Viding, Kumari, Blackwood, & Hodgins, 2013; Dolan & Park, 2002), we administered the Digits Backward, a 7-item measure from the Wechsler Digit Span Test. Participants hear a series of numbers at a rate of one number per second and are asked to repeat the numbers in reverse order. The first item of the Digits Backward task includes a pair of 3-digit series, and each item pair adds one digit from the previous item. The Digit Backwards is discontinued when individuals miss both series of digits in a given item pair. Scores are determined by summing the number of correct answers across all digit series and range from 0 to 14, with higher scores indicating a greater ability to examine, manipulate, and relay numerical information being held in working memory.

**Shipley Institute of Living Scale** (Zachary & Shipley, 1986). CD and APD are associated with lower Full-Scale IQs (Sánchez de Ribera, Kavish, Katz, & Boutwell, 2019). We administered the Shipley Institute of Living Scale to measure IQ and rule out the possibility that the diminished EV integration among individuals with more APD symptoms was driven by low IQ, The Shipley Institute of Living Scale includes two subtests: a 40-item multiple-choice vocabulary subtest in which participants select words that are synonyms of words provided and a 20-item pattern matching subtest in which participants write in responses to complete verbal and numerical patterns. Raw scores range from 0-80 and are used with conversion tables to find age-corrected t-scores and estimates of WAIS-R Full Scale IQ Scores. Higher scores indicate higher levels of general intelligence.

**Addiction Severity Index** (ASI; Leonhard, Mulvey, Gastfriend, & Shwartz, 2000). Individuals with more antisocial symptoms are likely to have more chronic patterns of heavy substance use (Brennan, Stuppy-Sullivan, Brazil, & Baskin-Sommers, 2017). In order to measure and later control for the chronicity of substance use, interviewers asked participants about their use of specific substances, including alcohol, cannabis, cocaine/crack, methamphetamines, other amphetamines, heroin, other opioids, hallucinogens, inhalants, nicotine, and other drugs. Interviewers recorded whether participants had ever tried a substance, the age at which participants first used the substance, and whether participants had regularly used the substance (three or more times per week). For participants who reported using a substance regularly, interviewers recorded age(s) when regular use started and ended to quantify the total number of years of regular use for each substance. The sum of years of regular use across substances provided an estimate of the chronicity of cumulative use.

**Self-Report Psychopathy-III** (SRP-III; Paulhus, Hemphill, & Hare, 2012). Antisocial behavior is a heterogeneous construct. Although most individuals who meet criteria for CD and APD symptoms reflect an externalizing antisocial subtype, a subset of these individuals reflect a psychopathic subtype, which is associated with distinct cognitive-affective dysfunctions from the externalizing subtype (Baskin-Sommers, Curtin, & Newman, 2015). We used the Self-Report Psychopathy-III (SRP-III) scale to measure psychopathic traits. The SRP-III is a 64-item self-report questionnaire that is intended to measure features (e.g., criminal tendencies, erratic lifestyle, interpersonal manipulation, and callous affect) of psychopathy. Items are scored on a 5-point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly). SRP-III total scores are sensitive to aspects of behavior that are common to multiple antisocial subtypes (e.g.,

criminal behavior, sensation seeking, impulsivity) and unique to psychopathy (e.g., interpersonal manipulation, shallow affect).

**Multidimensional Personality Questionnaire - Brief Form** (MPQ-BF; Patrick, Curtin, & Tellegen, 2002, Antisocial individuals are not only prone to negative affective states, but they also have a tendency towards negative emotionality (NEM), experiencing negative affect at a trait level (Baskin-Sommers, Curtin, & Newman, 2015; Patrick, Curtin, & Tellegen, 2002). To separate effects of negative affective states from trait NEM in our moderator analysis of negative affect, we administered the MPQ-BF, a 155-item self-report questionnaire that assesses personality traits across the lifespan. NEM is one of three orthogonal higher-order factors, encompassing stress reactivity, alienation, and aggression (Baskin-Sommers et al., 2015; Patrick et al., 2002). Higher NEM scores indicate higher levels of negative emotions at a trait level.

**Power Analysis**

Previous studies using the EEfRT detected three-way interactions among task variables and individual difference variables using sample sizes of 35-98 participants (Barch et al., 2014; Treadway et al., 2012; Treadway et al., 2009). Using GLIMMPSE Statistical Software (Kreidler et al., 2013), we opted to solve for sample size with a desired power of 80%. In specifying the study model, we included individual difference variables (APD symptoms, age, and a moderator variable) as predictors and expected value (EV) as a repeated-measures variable, and we selected the option to control for a single normally distributed covariate (to account for Trial effects; the GLIMMPSE statistical package does not include an option to control for two covariates). We selected the default study design for relatively equal group sizes and identified Choice (easy vs. hard) as the response variable. Given our interest in individual differences impacting APD symptoms and EV use, we specified an interaction hypothesis to include all pairwise

comparisons across two predictor variables (e.g., SUD severity and delay discount rate) and EV. We used the Hotelling-Lawley Trace statistical test option and specified a Type 1 Error rate of 0.05. When inputting predicted means, we specified mean differences of 15% and 20% hard choices as a function of EV, SUD severity, and a moderator variable, based on previous studies finding 3-way interactions among an EefRT task variable, a diagnostic variable, and an individual difference variable (Barch et al., 2014; Treadway et al., 2012; Treadway et al., 2009). To account for variability in study variables, for EV, we specified a base correlation of 0.3 and a decay rate of 0.05, and we requested that the LEAR model correlation matrix be computed with scaled spacing values. For Choice, we specified a standard deviation of 0.05 for each response. For the covariate (Trial), we specified an expected SD of 15 trials and a correlation of -0.25 with EV. Lastly, in options, we selected to use a Quantile method and specified a Quantile value of 0.5. Based on predicted mean differences of 15% and 20% hard task choices, calculated sample sizes were estimated at 80 and 48 participants, respectively. We recruited 94 participants to power our study to base the sample size on a more conservative effect size estimate and account for potential data loss.

## Supplemental Results

### Differentiating willingness to expend effort from ability to execute effortful tasks, neuropsychological functioning, chronic substance use, and psychopathic traits

To ensure that findings related to EV integration and antisociality were driven by an individual's willingness to exert effort to obtain rewards and not their ability to complete effortful tasks, we examined the completion rate across all trials for each subject. We found that all subjects completed between 80%-100% of trials. Additionally, we included the percentage of

successful trials with successful task completion as a covariate in the main analysis. The interaction between EV and APD symptoms remained significant ($p < 0.001$, 95% CI = -.264, -.093). Therefore, the association between diminished integration of EV and antisociality appeared independent from the ability to complete the button press tasks.

We also considered the possibility that the decreased willingness to exert effort in response to EV among individuals with more APD symptoms could reflect a decreased desire to spend time considering the effort-based decisions themselves. To examine whether individuals with higher levels of antisociality were less likely to use EV information when accounting for time spent to consider effortful tasks, we included choice reaction time as a covariate in the main analysis. The interaction between EV and APD symptoms remained significant ($p < 0.001$, 95% CI = -.253, -.083), indicating that the association between diminished integration of EV antisociality was independent from the time spent considering effort-based decisions.

It was also possible that the diminished EV integration among individuals with more APD symptoms actually reflected reduced fluency with numbers, diminished working memory ability, or low IQ. To examine whether the EV-by-APD symptom interaction remained when controlling for individual differences in these cognitive abilities, we reran the mixed effects model that included EV, APD symptoms, and the EV-by-APD symptoms interaction three separate times to include scores from neuropsychological measures of numeracy, working memory, and IQ. Results from these models indicated that the EV-by-APD symptoms interaction remained when controlling for measures of numeracy, working memory, and Full Scale IQ (EV-by-APD symptoms interaction controlling for Trails B: $p < 0.001$**,** 95% CI = -.259, -.089; EV-by-APD symptoms interaction controlling for Digits Backwards: $p < 0.001$, 95% CI =

-.260, -.089; EV-by-APD symptoms interaction controlling for IQ: $p < 0.001$, 95% CI = -.259, -.089).

Another factor that could influence effort-based decision-making was chronic substance use. We examined whether the EV-by-APD symptom interaction remained when controlling for years of regular substance use by including ASI total scores as a covariate. The result from this model indicated that the EV-by-APD symptoms effect appeared independent from chronic substance use (EV-by-APD symptom effect: $p < 0.001$, 95% CI = -.260, -.090). Individuals with more APD symptoms showed diminished use of EV, even controlling for the impact of chronic substance use.

Lastly, we included a model of EV use that included psychopathic traits as a covariate and found that the diminished EV use associated with antisociality was also independent of psychopathic traits (EV-by-APD symptoms interaction: $p < 0.001$, 95% CI = -.260, -.089). Thus, the diminished use of EV associated with antisociality appears to reflect an association between effort-based decision-making and the externalizing subtype of antisociality.

**Individuals with more APD symptoms show diminished cost-benefit integration during effort-based choice independent of current major depressive disorder symptoms and trait negative emotionality**

Current MDD symptom counts were added as a covariate to the logistic regression model with EV, APD symptoms., and the EV-by-APD symptoms interaction. Controlling for current MDD symptoms, the two-way interaction between EV and APD symptoms remained significant ($p < 0.001$, 95% CI = -0.259, -0.089). Therefore, the association between diminished integration of EV and antisociality appeared independent from major depressive symptoms.

To examine whether the EV-by-APD symptoms-by-negative affect interaction remained when controlling for individual differences in NEM at a trait level, we ran a separate model including NEM scores from the MPQ-BF as a covariate. The result from this model indicated that the diminished use of EV associated with negative affective and antisociality appeared independent from more trait-level individual differences in negative emotionality (B = -0.281, SE = 0.141, 95% C.I. = -0.558, -0.004, z = -1.99, $p$ = 0.046).

**Individuals with more APD symptoms and higher trait negative emotionality show diminished cost-benefit integration during effort-based choice**

We also examined whether negative emotionality (NEM), like negative affect, would moderate the relationship between EV and APD symptoms. We included EV, APD symptoms, NEM, and all two- and three-way interaction terms as continuous fixed-effect predictors. There was a significant EV-by-APD symptoms-by-NEM interaction (B = -0.012, SE = 0.003, 95% C.I. = 0.006, 0.017, z = -4.38, $p < 0.001$) whereby individuals with higher APD symptoms showed diminished EV use, whether they had higher or lower NEM. By contrast, individuals with lower APD symptoms showed diminished EV use if they had higher NEM and intact EV use if they had lower NEM (see Figure S1). Therefore, while the three-way interactions for both NEM and NA suggested that APD symptoms were associated with EV integration deficits, they reflect different relationships among NEM, NA, and APD symptoms in the context of effort-based decision making. Whereas the NEM interaction suggested an integration deficit that was common across individuals with higher levels of APD symptoms, higher levels of NEM, or their combination, the NA interaction suggested an integration deficit that was specific to a particular

subset of individuals who had more APD symptoms and were concurrently experiencing high levels of negative affect.

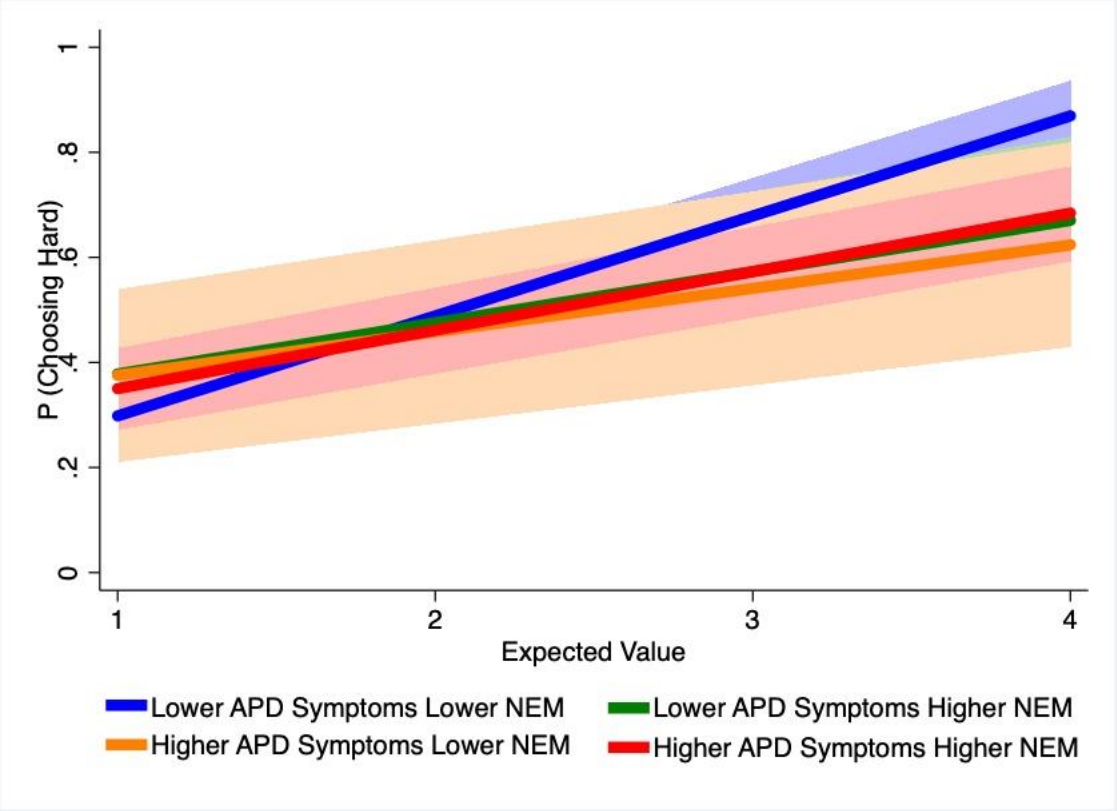


**Figure S1. Expected Value (EV) by APD symptoms and trait negative emotionality (NEM).** Compared to individuals with lower APD symptoms and lower NEM, individuals with higher APD symptoms and higher NEM were less likely to use EV to modulate their effort expenditure. Lines represent ±1 *SD* from the mean. Shading around lines represents 95% confidence intervals for point estimates.

## Delay discounting is associated with diminished use of probability, but not reward magnitude, during effort-based choice

To explore whether delay-discounting-linked differences in the use of EV information during effort-based choice reflected blunted sensitivity to one or both of the decision variables used to calculate EV, we completed a follow-up analysis to explore the two-way interactions between each decision variable and delay discounting. When we examined a model including the main effects of reward magnitude, probability, and ln*k* as well as the two-way interactions between the two task variables and ln*k* as continuous fixed-effect predictors, only sensitivity to

probability was blunted in individuals with higher levels of delay discounting (probability-by-ln*k*: B = -0.486, SE = 0.104, 95% C.I. = -0.690, -0.282, z = -4.67, $p < 0.001$). By contrast, sensitivity to reward magnitude of the hard task was intact among these individuals (reward magnitude-by-ln*k*: B = -0.022, SE = 0.036, 95% C.I. = -0.092, -0.048, z = -0.62, $p = .535$). These findings suggest that the relationship between delay discounting and effort-based decision-making may be specific to the use of probability information, rather than a deficit integrating multiple decision variables (i.e., reward and probability).

**Supplemental Table S1**

Other Psychiatric Diagnoses Present in Current Sample

| Diagnosis | Number of Participants Meeting Criteria |
|---|---|
| Major Depressive Disorder | 26 |
| Bipolar Disorder | 0 |
| Panic Disorder | 5 |
| Agoraphobia | 2 |
| Social Anxiety Disorder | 3 |
| Specific Phobia | 2 |
| Generalized Anxiety Disorder (Current/Past) | 3/0 |
| Obsessive-Compulsive Disorder | 3 |
| Bulimia Nervosa | 0 |
| Binge Eating Disorder | 0 |
| Attention Deficit/Hyperactivity Disorder | 2 |
| Acute Stress Disorder | 0 |
| Posttraumatic Stress Disorder | 14 |
| Alcohol Use Disorder Past Diagnosis | 38 |
| Alcohol Use Disorder Current Diagnosis | 2 |
| Sedative Use Disorder Past Diagnosis | 3 |
| Sedative Use Disorder Current Diagnosis | 2 |
| Cannabis Use Disorder Past Diagnosis | 29 |
| Cannabis Use Disorder Current Diagnosis | 17 |
| Stimulant Use Disorder Past Diagnosis | 16 |
| Stimulant Use Disorder Current Diagnosis | 6 |
| Opioid Use Disorder Past Diagnosis | 13 |
| Opioid Use Disorder Current Diagnosis | 3 |
| Inhalant Use Disorder Past Diagnosis | 0 |
| Inhalant Use Disorder Current Diagnosis | 0 |
| PCP Use Disorder Past Diagnosis | 1 |
| PCP Use Disorder Current Diagnosis | 0 |
| Hallucinogen Use Disorder Past Diagnosis | 4 |
| Hallucinogen Use Disorder Current Diagnosis | 1 |
| Other Use Disorder Past Diagnosis | 0 |
| Other Use Disorder Current Diagnosis | 0 |

**Supplemental References: Study 2**

Allen, D. N., & Haderlie, M. M. (2010). Trail-Making Test. *The Corsini Encyclopedia of Psychology*, 1-1.

Barch, D. M., Treadway, M. T., & Schoen, N. (2014). Effort, anhedonia, and function in schizophrenia: reduced effort allocation predicts amotivation and functional impairment. *Journal of Abnormal Psychology, 123*(2), 387.

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2015). Altering the cognitive-affective dysfunctions of psychopathic and externalizing offender subtypes with cognitive remediation. *Clin Psychol Sci, 3*(1), 45-57. doi:10.1177/2167702614560744

Brennan, G. M., Stuppy-Sullivan, A. M., Brazil, I. A., & Baskin-Sommers, A. R. (2017). Differentiating patterns of substance misuse by subtypes of antisocial traits in male offenders. *The Journal of Forensic Psychiatry & Psychology, 28*(3), 341-356.

De Brito, S. A., Viding, E., Kumari, V., Blackwood, N., & Hodgins, S. (2013). Cool and hot executive function impairments in violent offenders with antisocial personality disorder with and without psychopathy. *PLoS One, 8*(6).

Dolan, M., & Park, I. (2002). The neuropsychology of antisocial personality disorder. *Psychological medicine, 32*(3), 417-427.

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). Structured clinical interview for DSM-5—Research version (SCID-5 for DSM-5, research version; SCID-5-RV). *Arlington, VA: American Psychiatric Association*.

Khan, A. A., Jacobson, K. C., Gardner, C. O., Prescott, C. A., & Kendler, K. S. (2005). Personality and comorbidity of common psychiatric disorders. *The British Journal of Psychiatry, 186*(3), 190-196.

Kreidler, S. M., Muller, K. E., Grunwald, G. K., Ringham, B. M., Coker-Dukowitz, Z. T., Sakhadeo, U. R., . . . Glueck, D. H. (2013). GLIMMPSE: online power computation for linear models with and without a baseline covariate. *Journal of statistical software, 54*(10).

Leonhard, C., Mulvey, K., Gastfriend, D. R., & Shwartz, M. (2000). The Addiction Severity Index: A field study of internal consistency and validity. *Journal of Substance Abuse Treatment, 18*(2), 129-135. doi:10.1016/S0740-5472(99)00025-2

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment, 14*(2), 150.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and motor skills*, *8*(3), 271-276.

Sánchez de Ribera, O., Kavish, N., Katz, I. M., & Boutwell, B. B. (2019). Untangling intelligence, psychopathy, antisocial personality disorder, and conduct problems: A meta-analytic review. *European Journal of Personality*, *33*(5), 529-564.

Treadway, M. T., Bossaller, N. A., Shelton, R. C., & Zald, D. H. (2012). Effort-based decision-making in major depressive disorder: a translational model of motivational anhedonia. *Journal of Abnormal Psychology, 121*(3), 553.

Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E., & Zald, D. H. (2009). Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLoS One, 4*(8), e6598.

Wechsler, D. (1945). A standardized memory scale for clinical use. *The Journal of Psychology, 19*(1), 87-95.

Zachary, R. A., & Shipley, W. C. (1986). *Shipley institute of living scale: Revised manual*: WPS, Western Psychological Services.

# Appendix C: Study 3 Supplemental Material

## Supplementary Results

### Training Session Data Inclusion

We removed one participant's data for all 8 sessions due to notes that he fell asleep during multiple tasks across the majority of sessions he attended. For the Stop Signal Task, we followed guidelines outlined by Congdon and colleagues (2012) and removed session data for neutral and reward blocks when the probability of responding on stop signal trials was lower than 0.25 or higher than .75, when individuals omitted responses on over 40% of go trials, and when Stop Signal Task Reaction Time (SSRT) estimates were negative or less than 50 ms. For the decision-making task, we followed guidelines by Levy and colleagues (2010) and removed sessions on which individuals selected the lottery option on more than one catch trial. Additionally, for alpha and beta parameters, we followed cut-offs used by Konova and colleagues (2020), removing session data where behavior showed poor model fit (adjusted $r^2 <$ .1) and where values of the ambiguity tolerance parameter β were outside the classically interpretable range (-3 to 3). For Text Twist, participants appeared to understand the goal of the task, with participants completing words on each session across a normal distribution.

With these exclusions, we retained 83% of Stop Signal Task data for neutral trials, 85% of Stop Signal Task data for reward trials, 79% of decision-making data for the percent risk variable, 86% of decision-making data for the percent ambiguity data, 76% of decision-making data for the α parameter, 77% of decision-making data for the β parameter, and 93% of PASAT-C data for the persistence variable, compared to 100% of Text Twist data. Among participants completing CogTrain training, data for a given session was more likely to be excluded for

participants with low IQ, $r_s(115) = -.414$, $p < .001$ but was unrelated to a participant's level of externalizing as measured by the ESI-Brief total score, $r_s(118) = .096$, $p = .297$.

**Pre-post change by training condition using multiple imputation**

To avoid potential bias in the assessment of CogTrain effects because of missing data, we re-ran our pre-post change analyses for the cognitive-affective assessment and real-world behavior outcome variables with imputed values using SPSS Multiple Imputation. Missing posttreatment Digit Span Backwards, Breath Holding, Delay Discounting, substance use calendar, and TAU Session Attendance values were imputed in 26 data sets, based on recommendations to have as many imputations as the percentage of missing data for key outcome measures (Bodner, 2008). We included pretreatment measures of these variables, training condition, total number of training sessions, IQ scores, and age as predictors in the imputation model, selected the Automatic method, and specified constraints based on possible minimum and maximum values for outcome measures.

We completed multiple imputation analyses by pooling *F*-tests based on procedures set forth by van Ginkel and Kroonenberg (2014). For each analysis, we ran mixed effects models comparable to repeated measures ANOVA analyses, including effect-coded variables for timepoint and training condition, timepoint: (-1 [pretreatment], 1 [posttreatment]; training condition: -1 [Control], 1 [CogTrain]) as fixed-effect predictors. We also included the product of both effect coded variables to represent the two-way interaction between condition and timepoint. IQ (z-scored) was included as a continuous fixed effect predictor for the analysis of working memory, and age (z-scored) was included as a continuous fixed effect predictor for the analysis of primary drug substance use frequency. We also re-ran separate models including

pretreatment distress tolerance, trait anxiety, perceived stress, and change and motivation readiness as continuous covariates.

***Working Memory***

Multiple imputation results for working memory mirrored the complete-case only results, with a significant two-way interaction between timepoint and training condition, B = 0.514, SE = 0.214, 95% CI = 0.086, 0.942, t = 2.40, $p = 0.023$. Results from the covariate analyses also were similar to the complete-case results, with the two-way interaction between timepoint and training condition remaining significant when controlling for the influences of distress tolerance, trait anxiety, perceived stress, and change and motivation readiness.

***Primary Drug Substance Use Frequency***

For primary drug substance use frequency data, multiple imputation results differed from the complete-case results. With imputed data, the two-way interaction between timepoint and training condition was not significant, B = -0.054, SE = 0.080, 95% CI = -0.213, 0.106, $t = -0.67$, $p = 0.507$, suggesting that when individuals who provided substance use calendar data at pre-treatment only were included in the analysis, the cognitive remediation training package was not associated with pre-post reductions in substance use. To investigate why results with imputed data were different from results using complete-case data, we examined substance use patterns among individuals by training condition, posttreatment attendance, and pretreatment primary drug substance use frequency.

The majority of individuals who did not provide substance use data at posttreatment reported 28 days of abstinence during the pretreatment window (five of seven individuals in CogTrain, along with the only individual in the Control training condition who did not provide posttreatment data). To determine the likely imputed values for these individuals, we examined

data for individuals who were abstinent at pretreatment but provided data at posttreatment. We found that the majority of abstinent individuals (83% of participants in each group) who returned for posttreatment assessments remained abstinent at posttreatment. For individuals who had used their primary drug in the 28-day pretreatment window, posttreatment substance use frequency appeared to depend on the training condition they were assigned. The majority of the non-abstinent individuals in CogTrain (77%) reduced their substance use at posttreatment, compared to only 14% of individuals in the active control condition. Additionally, while only 22% of individuals in CogTrain increased their use in the 28 days of training, increased use was the modal response for individuals in the active control training (57% increased their substance use in the 28 days of training), and an additional 29% of active control participants made no changes to their substance use.

Overall, participants who were not represented in the complete-case analyses were most often individuals who had been abstinent at pretreatment and who were likely to have remained abstinent at posttreatment. Representing this possible trajectory (i.e., no change) in the multiple imputation analysis decreased the overall reductions in substance use observed among individuals in CogTrain, despite no change being the ideal outcome for individuals with this presentation. Data from the complete case analysis indicated that abstinent individuals were likely to remain abstinent in either training condition. However, among those who were actively using their primary drug at the pretreatment session, cognitive remediation training was most often associated with decreases in substance use (the ideal outcome), followed by no changes in use, with increases in use being the least common outcome for those receiving targeted training. By contrast, active control training was most often associated with increases in substance use,

followed by no changes in use, with decreases in use (the ideal outcome) being the least common outcome without targeted training.

***TAU session attendance***

TAU session attendance results using multiple imputation were consistent with the complete case results with one important exception. For the main analysis, there remained a significant two-way interaction between timepoint and training condition, B = 1.05, SE = 0.462, 95% CI = 0.130, 1.977, t = 2.28, $p$ = 0.028. Also consistent with the complete case results, the two-way interaction between timepoint and training condition remained significant when controlling for the influences of trait anxiety, perceived stress, and change and motivation readiness. However, diverging from the complete case results, the two-way interaction between timepoint and training condition was significant when controlling for the influence of pretreatment distress tolerance, B = 1.05, SE = 0.486, 95% CI = 0.130, 1.977, t = 2.28, $p$ = 0.028. This suggests that reduced statistical power in the complete case results may have decreased our ability to detect CogTrain-related increases in TAU session attendance when including an additional covariate. Our analysis that includes missing posttreatment data suggests that the observed increase in TAU session attendance among individuals in the CogTrain condition may in fact be attributable to completing targeted training, rather than being driven by the larger number of low distress tolerance individuals randomized into the CogTrain condition.

**Durability of effects by training condition**

We explored whether improvements on working memory, primary drug substance use frequency, and TAU session attendance among participants in the CogTrain condition persisted at the follow-up session one month after the posttreatment session, when individuals were no longer completing training tasks. For each of these analyses, we conducted a repeated measures

ANOVA with timepoint (pretreatment vs. posttreatment vs. 1-month follow-up) included as a within-subjects factor and training condition (CogTrain vs. Control) included as a between-subjects factor. For analyses of working memory, full scale IQ (*z*-scored) was again included as a continuous covariate, and for the analysis of primary substance use, age (*z*-scored) was included as a continuous covariate.

For working memory, there was a significant two-way interaction between timepoint and training condition when follow-up data was included, $F(2, 46) = 3.436$, $p = .041$, $\eta^2 = .13$, suggesting that the effect of timepoint was moderated by training condition. This significant interaction was decomposed using two orthogonal (Helmert) interaction contrasts designed to identify whether training condition moderated working memory change between pretreatment and later timepoints, and whether training condition moderated working memory change between posttreatment and follow-up. The first interaction contrast indicated training condition impacted working memory change from pretreatment to other timepoints, $F(1, 23) = 4.354$, $p = .048$, $\eta^2 = .16$, while the second interaction contrast indicated training condition did not impact change between posttreatment and the follow-up appointment, $F(1, 23) = 2.114$, $p = .160$, $\eta^2 = .08$. Follow-up simple-effects tests indicated that between pretreatment and posttreatment sessions, working memory significantly improved for individuals in the CogTrain condition (Mean Difference = 1.99, *SE* = 0.74, $p = 0.013$, 95% CI = [0.47, 3.52]), but did not significantly change for individuals in the active control condition (Mean Difference = -0.439, *SE* = 0.58, $p = 0.466$, 95% CI = [-1.97, 0.18]). By contrast, between posttreatment and follow-up sessions, working memory did not change significantly for either group (CogTrain: Mean Difference = -0.33, SE = 0.66, p = 0.624, 95% CI = [-1.69, 1.03]; active control: Mean Difference = 0.89, SE = 0.74, p = 0.013, 95% CI = [0.47, 3.52]), suggesting that pre-post change in working memory

among individuals in the CogTrain condition was sustained at the one-month follow-up, whereas working memory did not change significantly across any of the three timepoints for individuals in the Control condition.

For primary drug substance use frequency, the two-way interaction between timepoint and training condition was not significant when follow-up data was included, $F(1.72, 39.63)$[13] = 2.758, $p = .083$, $\eta^2 = .11$. Similarly, in the analysis for TAU session attendance, the two-way interaction between timepoint and training condition was not significant with the inclusion of data from follow-up $F(2, 60) = 1.255$, $p = .292$, $\eta^2 = .04$, These results suggest that pre-post change in primary drug substance use frequency and TAU session attendance did not differ by training condition when examining change across the full course of the study.

Notably, data at the one-month follow-up session was especially limited due to attrition, particularly in the CogTrain condition, with only 55% of the participants in the CogTrain condition compared to 83% of the participants in the active control training condition providing data at the one-month follow-up session. Therefore, significant effects for durability of change in working memory, and nonsignificant effects for change in primary drug substance use frequency and TAU session attendance should be interpreted with caution. It will be important for future studies of cognitive remediation training in antisocial individuals to recruit larger sample sizes, follow participants across multiple timepoints, and make special efforts to reduce attrition. Such efforts can help determine if training effects are durable and, if not, how booster sessions might be scheduled to reinforce progress over longer time periods optimally.

[13] Mauchly's test indicated that the assumption of sphericity had been violated for this effect, $\chi^2(2) = 8.87$; therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .86$).

## Supplemental References: Study 3

Bodner, T. E. (2008). What Improves with Increased Missing Data Imputations? *Structural equation modeling, 15*, 651-675.

Congdon, E., Mumford, J. A., Cohen, J. R., Galvan, A., Canli, T., & Poldrack, R. A. (2012). Measurement and reliability of response inhibition. *Frontiers in Psychology, 3*, 37.

Konova, A. B., Lopez-Guzman, S., Urmanche, A., Ross, S., Louie, K., Rotrosen, J., & Glimcher, P. W. (2020). Computational markers of risky decision-making for identification of temporal windows of vulnerability to opioid use in a real-world clinical setting. *JAMA Psychiatry, 77*(4), 368-377.

Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *J Neurophysiol, 103*(2), 1036-1047. doi:10.1152/jn.00853.2009

van Ginkel, J. R., & Kroonenberg, P. M. (2014). Analysis of variance of multiply imputed data. *Multivariate behavioral research, 49*(1), 78-91.