UCLA
UCLA Electronic Theses and Dissertations

Title

Benefit, Burdens and Brain Correlates of Heightened Reward Sensitivity in Adolescence

Permalink

https://escholarship.org/uc/item/5p59v40c

Author

Stolyarova, Alexandra

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Benefit, Burdens and Brain Correlates of Heightened Reward Sensitivity in Adolescence

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Psychology

by

Alexandra Stolyarova

2020

ABSTRACT OF THE DISSERTATION

Benefit, Burdens and Brain Correlates of Heightened Reward Sensitivity in Adolescence

by

Alexandra Stolyarova

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2020

Professor Alicia Izquierdo Edler, Chair

Increased exploration, risk-taking and reward-seeking are the hallmarks of adolescence. These and related behaviors prepare the young to transition from the parent's nest to independent living. Adolescence is also a period of heightened structural and functional brain reorganization, particularly within the mesocorticolimbic dopamine system, frontal cortex, striatum and amygdala - an interconnected network of brain regions that supports reward-directed behavior. The goal of this dissertation to further our understanding of ways in which the adolescent brain interacts with rewards. In **Chapter 2**, I present the results of a set of experiments demonstrating that adolescent rats do not differ from adults in a simple form of stimulus-reward learning but are willing to invest greater effort to obtain larger rewards. While previous studies have focused on the role of neurodevelopmental changes in the dopamine system and within the striatum in heightened reward seeking in adolescence, our data suggest that synaptic remodeling within the frontal cortex and amygdala may also contribute to enhanced reward sensitivity. **Chapter 3** explores the long-term consequences of prescription drug exposure during the adolescent period of heightened reward

sensitivity in rats. The data suggest that adolescent exposure to both fluoxetine and methylphenidate impairs learning and cognitive flexibility in adulthood in male rats long after the drug administration has been terminated. Adolescent methylphenidate exposure has the direst consequences, impairing both the initial learning and reversal of reward contingencies. The data also reveal sex differences: while females take longer to learn the task, they are also less vulnerable to the negative effects of drug exposure. In **Chapter 5**, I compare adult and adolescent humans in their approaches to solving the *credit assignment problem* (i.e., the problem of discovering which choices are responsible for rewards, introduced in **Chapter 4**). The data provide preliminary evidence for enhanced contingent credit assignment in adolescence. While adults integrate information about their previous decisions and past outcomes to guide their subsequent choices, adolescents are more narrowly focused on the most recent choice-reward history. In **Chapter 6**, I discuss the implications of the present work and offer cautious advice on drug abuse prevention and improvements in pedagogical practice.

The dissertation of Alexandra Stolyarova is approved.

Michele Basso

Dean Buonomano

Adriana Galván

Hakwan Lau

Alicia Izquierdo Edler, Committee Chair

University of California, Los Angeles

2020

To Alicia,

without whom there simply would not be Alex, a scientist.

Table of Contents

List of figures

Lastly, I would like to acknowledge my family. My mother, Irina Stolyarova, father, Alexander Stolyarov, and grandmother, Natalia Kolosunina, have instilled in me passion for knowledge and education and loved, encouraged and supported me at every step of the way. My in-laws, Kristi, George-father, George-brother, Stephanie, Harrison Mansolf, Josh Spohrer and Zelda and Cosmos Mansolf-Spohrer, I am deeply indebted to you for the unconditional acceptance into your family and for lifting my spirits during the difficult times. Without my husband, Maxwell Mansolf, this dissertation would have never been finished. Max, you have made me feel safe and your love and support got me though the darkest times. Finally, I would like to thank little Maxwell, Evan and Alex for their emotional support, encouragement, optimistic outlook, great sense of humor, and being so fuzzy and cute. Thank you all.

Permissions. Portions of this dissertation are adapted from previously published articles. The data are published in:

Stolyarova A, Izquierdo A (2015) Distinct patterns of outcome valuation and amygdala-prefrontal

cortex synaptic remodeling in adolescence and adulthood. *Front Neurosci* 9:115.

https://doi.org/10.3389/fnbeh.2015.00115

Portions of this manuscript are included in a modified form in Chapter 3. Novel statistical analyses

were conducted and results were reported in this dissertation.

Izquierdo A, Pozos H, De La Torre A, DeShields S, Cevallos J, Rodriguez J, Stolyarova A (2016)

Methylphenidate, fluoxetine, and methamphetamine in rat adolescence: Effects on learning

flexibility, emotion, and striatal dopamine D1 and D2 receptor expression in adulthood.

Behav Brain Res 308:104-114. https://doi.org/10.1016/j.bbr.2016.04.028

Portions of this manuscript are included in a modified form in Chapter 3. Novel statistical analyses

were conducted and results were reported in this dissertation. Additionally, the PSA-NCAM data

reported in Chapter 3 of this dissertation were not included in the original publication.

Stolyarova A (2018) Solving the credit assignment problem with the prefrontal cortex. *Front*

Neurosci 2:182. https://doi.org/10.3389/fnins.2018.00182

Portions of this manuscript are included in a modified form in Chapter 4.

Contributions of co-authors and non-authors. For work that has been previously

published, I would like to thank my co-authors. For Chapter 2, I would like to acknowledge Dr.

Alicia Izquierdo who contributed to writing of the original manuscript on which the chapter is

based. I would also like to thank Adrianna De La Torre and Amador Bugarin for their help in

collecting the operant learning data, as well as Dr. Andrew B. Thompson, Eric Harvey, and

Amandeep Kaur for help with tissue processing and ELISA assays reported in Chapter 2. Hilda

VITA

Honors and Awards

The Collegium of University Teaching Fellow	2019-2020
UCLA Dissertation Year Fellowship	2019-2020
Atamdede Fellowship	2019
Nancy M. Biram Research Fund in Life Sciences Award	2019
Shepherd Ivory Franz Distinguished Teaching Award	2018
Charles E. and Sue K. Young Graduate Student Fellowship	2018
Society for Neuroscience Trainee Professional Development Award	2016
UCLA Brain Research Institute Graduate Student Travel Award	2016, 2015
International Behavioral Neuroscience Society (IBNS) Travel Award	2016
The RIKEN Brain Science Institute Travel Award	2016
Marine Biological Laboratory Travel Award and Fellowship	2016
UCLA Graduate Summer Research Mentorship Fellowship	2016
UCLA Distinguished University Fellowship	2015-2016

Publications

Aguirre CG, **Stolyarova A**, Das K, Kolli S, Marty V, Ray L, Spigelman I, Izquierdo A (2020) Sex-dependent effects of chronic intermittent voluntary alcohol consumption on attentional, not motivational, measures during probabilistic learning and reversal. *PLoS One* 15(6): e0234729.

Stolyarova A, Wikenheiser AM (2020) Can the VTA come out to play? Only when the mPFC's predictions go astray. *Neuron* 105(4):593-595.

Stolyarova A[#], Rakhshan M[#], Hart EE, O'Dell TJ, Peters MAK, Lau H, Soltani A, Izquierdo A (2019) Dissociable roles for anterior cingulate cortex and basolateral amygdala in decision confidence and learning under uncertainty. *Nature Communications* 10(1):4704. [#]*co-first*

Izquierdo A, Aguirre C, Hart EE, **Stolyarova A** (2019) Animal models of adaptive learning and decision making. *Methods in Molecular Biology: Psychiatric Disorders* (2nd Edition): 105-119.

Stolyarova A (2018) Solving the credit assignment problem with the prefrontal cortex. *Front Neurosci* 2:182.

Stolyarova A, Izquierdo A (2017) Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning and decision making under uncertainty. *Elife.* 6. pii: e27483.

Izquierdo A, Pozos H, De La Torre A, DeShields S, Cevallos J, Rodriguez J, **Stolyarova A** (2016) Methylphenidate, fluoxetine, and methamphetamine in rat adolescence: Effects on learning flexibility, emotion, and striatal dopamine D1 and D2 receptor expression in adulthood. *Behav Brain Res* 308:104-114.

Thompson AB, **Stolyarova A**, Ying Z, Zhuang Y, Gomez-Pinilla F, Izquierdo A (2015) Methamphetamine blocks exercise effects on Bdnf and Drd2 gene expression in frontal cortex and striatum. *Neuropharmacology* 99:658-64.

Stolyarova A, Izquierdo A (2015) Distinct patterns of outcome valuation and amygdala-prefrontal cortex synaptic remodeling in adolescence and adulthood. *Front Neurosci* 9: 115.

Ochoa G, **Stolyarova A**, Kaur A, Hart E, Bugarin A, Izquierdo A (2015) Post-training depletions of basolateral amygdala serotonin fail to disrupt discrimination, retention, or reversal learning. *Front Neurosci* 9: 155.

Stolyarova A, Thompson A, Barrientos RM, Izquierdo A (2015) Reductions in frontocortical cytokine levels are associated with long-lasting alterations in reward valuation after methamphetamine. *Neuropsychopharmacology* 40(5): 1234-42.

Stolyarova A, O'Dell SJ, Marshall JF, Izquierdo A (2014) Positive and negative feedback learning and associated dopamine and serotonin transporter binding after methamphetamine. *Behav Brain Res* 271: 195-202.

Teaching

Seminar: Psychology of (Ir)Rational Decision-Making: Learning to Make Better Choices and Outsmart Your Biases, UCLA (Instructor of Record): Winter 2020

Behavioral Neuroscience Laboratory, UCLA (Teaching Fellow): Spring 2020, 2017; Fall 2019, 2018; Summer 2018, 2017

Introduction to Psychology, UCLA (Instructor of Record): Summer 2019, 2018

Psychology of Human Behavior, UCLA Extensions (Instructor of Record): Summer 2019

Advanced Psychological Statistics: Regression, UCLA (Teaching Associate): Spring 2018

Advanced Psychological Statistics: B, UCLA (Teaching Associate): Winter 2018

Advanced Psychological Statistics: A, UCLA (Teaching Associate): Fall 2017

Comparative Psychobiology, UCLA (Teaching Associate): Summer 2017

From Molecules to Mind - Molecular and Developmental Neuroscience, UCLA (Teaching Assistant): Winter 2017

Introduction to Biopsychology, UCLA (Teaching Assistant): Fall 2016

Chapter 1: General introduction

Increased exploration, risk-taking and reward-seeking are the hallmarks of adolescence. These and related behaviors prepare the young to transition from the parent's nest to independent living: heightened risk- and reward-seeking, for example, ensure that the adolescent meets the nutritional needs of her rapidly growing body and establishes strong relationships with her peer group (van Duijvenvoorde et al., 2016; Doremus-Fitzwater et al., 2010; Steinberg et al., 2009; Urosevic et al., 2012; Andersen et al., 2002; Doremus-Fitzwater and Spear, 2016). The desire to procure rewards also encourages the adolescent to excel academically and in hobbies or sports (Telzer, 2016). Throughout human history, adolescent exploration, risk-taking and status-seeking have contributed to serendipitous innovation and rapid territorial and technological expansion, therefore also benefiting the group (Dahl et al., 2018). On the flip side, increased reward-seeking can predispose adolescents to harmful and dangerous behaviors, including reckless driving, unprotected sex and experimentation with drugs (Galván et al., 2007).

Adolescence is also a period of structural and functional brain reorganization. While many sensory functions, as, for example, famously demonstrated by Wiesel and Hubel (1963, 1965) for vision, and language undergo a period of augmented neural plasticity early in life, adolescence represents a 'second period of heightened malleability' (Steinberg, 2014; Fuhrmann et al., 2015). The mesocorticolimbic dopamine (DA) system, the frontal cortex, striatum and amygdala - an interconnected network of brain regions that supports reward-directed behavior and learning, in particular, is at the peak of structural and functional remodeling during adolescence (Andersen et al., 2000; Andersen, 2002; Tarazi and Baldessarini, 2000; Brenhouse et al., 2008; Benes et al., 2000; Naneix et al., 2012; Weickert et al., 2007; Pattwell et al., 2016; Arruda-Carvalho et al., 2017). In the following paragraphs of the introduction, I overview in greater detail the relationship

between neurodevelopmental changes in these brain regions and reward-guided behaviors in adolescence both in humans and in the rodent model.

The age of adolescence across species

In both humans and rats, the beginning and the end of the adolescent period are marked by a combination of biological, including neural, behavioral and environmental, including social, factors. In humans, the transition to adolescence begins with the onset of puberty by the age of 10 years in girls and 12 years in boys (Dahl et al., 2018). The end of adolescence has been more difficult to define in humans as it depends on a combination of biological, cognitive, affective, and social criteria. For example, the onset of adulthood in humans is usually associated with adopting certain social roles and being granted certain legal rights and responsibilities (Dahl et al., 2018). The ambiguity of the criteria has also led to fluctuations in the adolescent vs adult cut-off points across studies (e.g., range 17-24 years old), which may lead to difficulties in the synthesis of results. The rat adolescent period spans approximately 4.5 weeks from the post-natal day (PND) 28 to PND 60, with the last week corresponding to late adolescence/emerging adulthood (Doremus-Fitzwater and Spear, 2016). During adolescence, the body undergoes a period of sexual maturation and accelerated growth, accompanied by changes in sleep, circadian and hormonal regulation, and metabolism (Dahl et al., 2018). The hormonal changes initiate a cascade of events that affects brain development, particularly in areas involved in regulating cognitive and emotional processes, social behavior and reward-guided learning and decision-making (Dahl et al., 2018).

The adolescent brain seeks rewards

Reward seeking peaks in adolescence for both human and rodent species (van Duijvenvoorde et al., 2016; Doremus-Fitzwater et al., 2010; Steinberg et al., 2009; Urosevic et al., 2012; Andersen et al., 2002; Doremus-Fitzwater and Spear, 2016) as compared to both earlier childhood and later adulthood. Adolescents show a higher preference for sweet, tasty food and drinks and are more likely to bet on risky options associated with larger rewards, even if those options are less advantageous in the long run (Cauffman et al., 2010; Desor and Beauchamp, 1987; Friemel et al., 2010; Wilmouth and Spear, 2009; Doremus-Fitzwater and Spear, 2016). Compared to adults, adolescents are also more likely to prefer the immediately available reward to the more beneficial option available after a prolonged delay (Huang et al., 2017).

The developmental changes in an evolutionarily conserved network of densely interconnected brain regions have been linked to the characteristic increase in reward seeking during adolescence. Of particular relevance is the structural and functional reorganization within the mesocorticolimbic DA system, the frontal cortex, striatum and the amygdala. The D1-like and D2-like receptors in the ventral striatum have been shown to increase in density from childhood to mid-adolescence followed by pruning to the lower adulthood levels thereafter (Andersen et al., 2000; Andersen, 2002; Tarazi and Baldessarini, 2000). The glutamatergic inputs from the frontal cortex to the striatum also undergo remodeling, gradually increasing in number throughout the adolescent period (Brenhouse et al., 2008). These changes to the synaptic organization within the striatum likely contribute to the increase in reward seeking in adolescents, given that in adults the striatal DA has been linked to the invigoration of responding toward rewards (Berridge and Robinson, 1998; Robbins and Everitt, 2007; Doremus-Fitzwater and Spear, 2016). Furthermore, in adolescent humans the ventral striatum is more active in response to the receipt of rewards (Ernst

et al., 2005; Galván et al., 2006; Van Leijenhorst et al., 2010; Galván and McGlennen, 2013; Chein et al., 2010, Guyer et al., 2009).

Within the frontal cortex, the development of DA projections follows a delayed time course compared to the striatum: the density of DA synapses starts to increase just before the onset of adolescence and continues into early adulthood, and the expression of D1-like and D2-like receptors in the frontal cortex reaches its peak by late, rather than mid, adolescence followed by pruning into adulthood (Benes et al., 2000; Andersen et al., 2000; Andersen et al., 2002; Naneix et al., 2012; Weickert et al., 2007). The upregulation of D1 receptors, specifically within the frontal cortex, appears to underly many of the unique ways in which adolescents interact with rewards. An artificial overexpression of D1 receptors in the frontal cortex of adult rodents is sufficient to induce the preference for sweet solutions, increase impulsivity and potentiate drug responsivity (Sonntag et al., 2014).

The amygdala is another brain region critical for reward-guided behavior that undergoes substantial remodeling during adolescence (Wassum and Izquierdo, 2015; Walker at al., 2017). In particular, the number of glutamatergic synapses from the frontal cortex onto amygdala neurons peaks in early adolescence followed by gradual pruning into adulthood (Pattwell et al., 2016). Early adolescence is also characterized by synaptic strengthening of the frontal cortex-to-amygdala projections (Arruda-Carvalho et al., 2017).

The adolescent brain learns from reinforcement

The ability to increase the frequency of behavior in response to reward (i.e., learn from reinforcement) is established early in life: even 10-week-old infants can learn within 6 minutes to kick their feet to move a mobile hanging overhead (Rovee and Rovee, 1969; Nussenbaum and

Hartley, 2019). Across development, reinforcement learning becomes more complex and adaptable: the brain learns-to-learn from secondary and probabilistic rewards, from counterfactual information, and in dynamic environments. Adolescents do not differ from adults in their baseline learning rates or in their ability to respond to abrupt changes in reward contingencies (Javadi et al., 2014). However, adults appear to be better at calibrating their learning to reward statistics on the task compared to adolescents (Nussenbaum and Hartley, 2019; Decker et al., 2015; Master et al., 2019).

The DA system has long been thought to play a central role in reinforcement learning. Since the hallmark observation by Schultz and colleagues (Schultz 1998; Schultz et al., 1998), DA neurotransmission has been parsimoniously implicated in signaling unexpected outcomes of behaviors, termed reward prediction errors, across species and behavioral paradigms. Most DA cells in the midbrain increase their firing in response to the delivery of unexpected rewards and suppress their activity in response to the omission of predicted rewards. Such signals are thought to be used by downstream structures, including the striatum, amygdala, and frontal cortex to learn from the rewarding experience (or its omission).

The neural responses to reward prediction errors within the striatum on a probabilistic task have been shown to peak in amplitude during adolescence compared to adulthood (Cohen et al., 2010). Despite the larger responses to violation in reward prediction, however, adolescents do not appear to learn the tasks faster than adults. On probabilistic tasks, one must learn that not all rewards are informative: in fact, even the worst option will produce a reward on a small subset of trials. Given the observation that adolescents are not as efficient as adults at calibrating their learning to the task reward statistics, they may inappropriately learn to prefer the worse option from the salient positive feedback on the rare occasions when that option produces rewards.

Some studies also reported developmental differences in the relative rates of learning from positive vs negative feedback. However, the direction of the effect varied from study to study, with one report demonstrating an increased and another a decreased reliance on negative feedback in adolescents compared to adults (van den Bos et al., 2012; van den Bos et al., 2009; DePasque and Galván, 2017). Intriguingly, however, the prefrontal cortex and its connectivity with the striatum was associated with differences in learning from positive vs negative feedback in both studies.

Dissertation overview

The goal of this dissertation is to further our understanding of ways in which the adolescent brain interacts with rewards in its environment. In **Chapter 2**, I present the results of a set of experiments demonstrating that adolescent rats do not differ from adults in a simple form of stimulus-reward learning but show an enhanced motivation to invest effort to obtain larger rewards. These findings highlight the general pattern of reward sensitivity in adolescents, observed on tasks with risk, delay and effort requirements and across species. The data presented in Chapter 1 also provide preliminary evidence that synaptic remodeling within the frontal cortex and amygdala contributes to reward sensitivity in adolescence. **Chapter 3** explores the long-term behavioral and neural consequences of exposure to drugs frequently prescribed or misused/abused during the adolescent period of heightened reward sensitivity in rats. The data suggest that adolescent exposure to both fluoxetine and methylphenidate impairs learning and cognitive flexibility in adulthood in male rats, long after the drug administration has been terminated. The data also reveal sex differences: while females take longer to learn the task, they are also less vulnerable to the negative effects of adolescent drug exposure. Analyses of protein expression after learning reveal an upregulation of striatal D1 receptors in adulthood following adolescent

methamphetamine (in males) and methylphenidate (in females) exposure, an upregulation of striatal D2 receptors following adolescent methylphenidate (in females) exposure and higher levels of cortical PSA-NCAM expression in male, compared to female, animals. These findings suggest that developmental psychostimulant exposure may interact with reward experience to boost striatal D1 and D2 receptor expression in a sex-dependent manner, later in life. **Chapters 4 and 5** characterize the learning strategies adopted by adolescent humans on a probabilistic task. In particular, I compare adults and adolescents in their approaches to solving the *credit assignment problem* (i.e., the problem of discovering which choices are responsible for rewards). The data provide preliminary evidence for enhanced contingent credit assignment in adolescence: the general pattern of results suggests an enhanced sensitivity to contingent feedback and decreased sensitivity to non-contingent feedback in adolescents compared to adults. While adults integrate information about their previous decisions and past outcomes to guide their subsequent choices, adolescents are more narrowly focused on the most recent choice-reward history. In **Chapter 6**, I discuss the implications of the present work and offer cautious advice on drug abuse prevention and improvements in pedagogical practice.

References

Andersen, S. L., Thompson, A. P., Krenzel, E., Teicher, M. H. (2002). Pubertal changes in gonadal hormones do not underlie adolescent dopamine receptor overproduction. *Psychoneuroendocrinology* 27, 683–691.

Andersen, S. L., Thompson, A. T., Rutstein, M., Hostetter, J. C., Teicher, M. H. (2000). Dopamine receptor pruning in prefrontal cortex during the periadolescent period in rats. *Synapse* 37, 167–169.

Arruda-Carvalho, M., Wu, W. C., Cummings, K. A., Clem, R. L. (2017). Optogenetic examination of prefrontal-amygdala synaptic development. *J Neurosci* 37, 2976–2985.

Benes, F. M., Taylor, J. B., Cunningham, M. C. (2000). Convergence and plasticity of monoaminergic systems in the medial prefrontal cortex during the postnatal period: implications for the development of psychopathology. *Cereb Cortex* 10, 1014–1027.

Berridge, K. C., Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Brain Res Rev* 28, 309–369.

Brenhouse, H. C., Sonntag, K. C., Andersen, S. L. (2008). Transient D1 dopamine receptor expression on prefrontal cortex projection neurons: relationship to enhanced motivational salience of drug cues in adolescence. *J Neurosci* 28, 2375–2382.

Cauffman, E., Shulman, E. P., Steinberg, L., Claus, E., Banich, M. T., Graham, S., et al. (2010). Age differences in affective decision making as indexed by performance on the Iowa Gambling Task. *Dev Psychol* 46, 193–207.

Chein, J., Albert, D., O'Brien, L., Uckert, K., Steinberg, L. (2010). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Dev Sci* 14(2), F1–F10.

Cohen, J. R., Asarnow, R. F., Sabb, F. W., Bilder, R. M., Bookheimer, S. Y., Knowlton, B. J., et al. (2010). A unique adolescent response to reward prediction errors *Nature Neuroscience* 13(6), 669-671.

Dahl, R. E., Allen, N. B., Wilbrecht, L., Suleiman, A. B. (2018). Importance of investing in adolescence from a developmental science perspective. *Nature* 554(7693), 441-450.

Decker, J.H., Lourenco, F.S., Doll, B.B., Hartley, C.A. (2015). Experiential reward learning outweighs instruction prior to adulthood. *Cogn Affect Behav Neurosci* 15(2), 310–320.

DePasque, S., Galván, A. (2017). Frontostriatal development and probabilistic reinforcement learning during adolescence. *Neurobiol Learn Mem* 143, 1-7.

Desor, J. A., Beauchamp, G. K. (1987). Longitudinal changes in sweet preferences in humans. *Physiol Behav* 39, 639–641.

Doremus-Fitzwater, T. L., Spear, L. P. (2016). Reward-centricity and attenuated aversions: An adolescent phenotype emerging from studies in laboratory animals. *Neurosci Biobehav Rev* 70, 121–134.

Doremus-Fitzwater, T. L., Varlinskaya, E. I., Spear, L. P. (2010). Motivational systems in adolescence: possible implications for age differences in substance abuse and other risk-taking behaviors. *Brain Cogn* 72, 114–123.

Ernst, M., Nelson, E. E., Jazbec, S., McClure, E. B., Monk, C. S., Leibenluft, E., et al. (2005). Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. *Neuroimage* 25(4), 1279–1291.

Friemel, C. M., Spanagel, R., Schneider, M. (2010). Reward sensitivity for a palatable food reward peaks during pubertal developmental in rats. *Front Behav Neurosci* 4.

Fuhrmann, D., Knoll, L. J., Blakemore, S. J. (2015). Adolescence as a Sensitive Period of Brain Development. *Trends Cogn Sci* 19(10), 558-566.

Galván, A., McGlennen, K. (2013). Enhanced striatal sensitivity to aversive reinforcement in adolescents versus adults. *J Cogn Neurosci* 25, 284–296.

Galván, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., et al. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *J Neurosci* 26(25), 6885–6892.

Galván, A., Hare, T., Voss, H., Glover, G., Casey, B. J. (2007). Risk-taking and the adolescent brain: Who is at risk? *Developmental Science* 10(2), F8-F14

Guyer, A. E., McClure-Tone, E. B., Shiffrin, N. D., Pine, D. S., Nelson, E. E. (2009). Probing the neural correlates of anticipated peer evaluation in adolescence. *Child Dev* 80(4), 1000–1015.

Huang, Y., Hu, P., Li, X. (2017). Undervaluing delayed rewards explains adolescents' impulsivity in inter-temporal choice: an ERP study. *Sci Rep* 7, 42631.

Javadi, A. H., Schmidt, D. H. K., Smolka, M. N. (2014). Adolescents adapt more slowly than adults to varying reward contingencies. *J Cogn Neurosci* 26(12), 2670–2681.

Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., Collins, A. G. E. (2020). Distentangling the systems contributing to changes in learning during adolescence. *Dev Cogn Neurosci* 41:100732.

Naneix, F., Marchand, A. R., Di Scala, G., Pape, J. R., Coutureau, E. (2012). Parallel maturation of goal-directed behavior and dopaminergic systems during adolescence. *J Neurosci* 32, 16223–16232.

Nussenbaum, K., Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Dev Cogn Neurosci*. 40: 100733.

Pattwell, S. S., Liston, C., Jing, D., Ninan, I., Yang, R. R., Witztum, J., et al. (2016). Dynamic changes in neural circuitry during adolescence are associated with persistent attenuation of fear memories. *Nat Commun* 7:11475.

Robbins, T. W., Everitt, B. J. (2007). A role for mesencephalic dopamine in activation: commentary on Berridge (2006) *Psychopharmacology (Berl)* 191, 433–437.

Rovee, C. K., Rovee, D. T. (1969). Conjugate reinforcement of infant exploratory behavior. *J Exp Child Psychol* 8(1), 33–39.

Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. Journal of Neurophysiology 80(1), 1–27.

Schultz, W., Tremblay, L., Hollerman, J. R. (1998). Reward Prediction in Primate Basal Ganglia and Frontal Cortex. *Neuropharmacology* 37(4-5), 421–29.

Sonntag, K. C., Brenhouse, H. C., Freund, N., Thompson, B. S., Puhl, M., Andersen, S. L. (2014). Viral over-expression of D1 dopamine receptors in the prefrontal cortex increase high-risk behaviors in adults: comparison with adolescents. *Psychopharmacology (Berl)* 231, 1615–1626.

Steinberg, L. (2014). Age of Opportunity – Lessons from the New Science of Adolescence. *Houghton Mifflin Harcourt.*

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., Banich, M. (2009). Age differences in future orientation and delay discounting. *Child Dev* 80, 28–44.

Tarazi, F. I., Baldessarini, R. J. (2000). Comparative postnatal development of dopamine D(1), D(2) and D(4) receptors in rat forebrain. *Int J Dev Neurosci* 18, 29–37.

Telzer, E. H. (2016). Dopaminergic reward sensitivity can promote adolescent health: A new perspective on the mechanism of ventral striatum activation. *Dev Cogn Neurosci.* 17, 57-67.

Urosevic, S., Collins, P., Muetzel, R., Lim, K., Luciana, M. (2012). Longitudinal changes in behavioral approach system sensitivity and brain structures involved in reward processing during adolescence. *Dev Psychol* 48, 1488–1500.

van den Bos, W., Cohen, M. X., Kahnt, T., Crone, E. A. (2012). Striatum-Medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral Cortex* 22(6), 1247-1255

van den Bos, W., Güroğlu, B., van den Bulk, B. G., Rombouts, S. A. R. B., Crone, E. A. (2009). Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. *Frontiers in Human Neuroscience* 3, 52

van Duijvenvoorde, A. C, Peters, S., Braams, B. R., Crone, E. A. (2016). What motivates adolescents? Neural responses to rewards and their influence on adolescents' risk taking, learning, and cognitive control. *Neurosci Biobehav Rev* 70, 135-147

Van Leijenhorst, L., Moor, B. G., de Macks, Z. A. O., Rombouts, S. A., Westenberg, P. M., Crone, E. A. (2010). Adolescent risky decision-making: neurocognitive development of reward and control regions. *Neuroimage* 51(1), 345–355.

Walker, D. M., Bell, M. R., Flores, C., Gulley, J. M., Willing, J., Paul, M. J. (2017). Adolescence and Reward: Making Sense of Neural and Behavioral Changes Amid the Chaos. *J Neurosci* 37(45), 10855-10866.

Wassum, K. M., Izquierdo, A. (2015). The basolateral amygdala in reward learning and addiction. *Neurosci Biobehav Rev* 57, 271–283.

Weickert, C. S., Webster, M. J., Gondipalli, P., Rothmond, D., Fatula, R. J., Herman, M. M., et al. (2007). Postnatal alterations in dopaminergic markers in the human prefrontal cortex. *Neuroscience* 144, 1109–1119.

Wiesel, T. N., Hubel, D. H. (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J Neurophysiol* 28, 1029-1040.

Wiesel, T. N., Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J Neurophysiol* 26, 1003-1017

Wilmouth, C. E., Spear, L. P. (2009). Hedonic sensitivity in adolescent and adult rats: taste reactivity and voluntary sucrose consumption. *Pharmacol Biochem Behav* 92, 566–573.

Chapter 2: Reward valuation and synaptic remodeling in adolescence and adulthood

Abstract

Adolescent behavior is characterized by increased risk-taking, reward- and novelty-seeking, as well as an augmented need for social and environmental stimulation. This behavioral phenotype may result from alterations in outcome valuation or learning about rewards. In the present set of experiments, we directly compared adult and adolescent animals on tasks measuring both of these processes. Additionally, we examined developmental differences in dopamine D1-like receptor (D1R), dopamine D2-like receptor (D2R), and polysialylated neural cell adhesion molecule (PSA-NCAM) expression (a molecule associated with synaptic remodeling) in animals that were trained on an effortful reward valuation task, given that these proteins play an important role in the functional development of the amygdala-frontocortical (FC) circuitry and the mesocorticolimbic dopamine system. We found that adolescent animals were not different from adults in a simple form of appetitive associative learning but showed an enhanced motivation to invest effort to obtain larger rewards. There were no differences in D2 receptor expression, but D1 receptor expression was significantly reduced in the striata of animals that had experiences with reward learning during adolescence compared to animals that went through the same experiences in adulthood. We observed increased levels of PSA-NCAM expression in both the FC and amygdala of late adolescents compared to adults that were previously trained on an effortful reward valuation task. Increased levels of PSA-NCAM expression in adolescents may index increased structural plasticity and represent a neural correlate of a reward sensitive endophenotype.

Introduction

Adolescence is a critical period during which animals learn to predict future states of their habitat depending on current experiences and acquire life strategies that are likely to promote survival and reproductive success later in life. The fitness is increased if the phenotype that developed in early life is matched to the predicted environment (Gluckman et al., 2007), and if an animal can adequately cope with the environmental uncertainty and reward availability (McNamara et al., 2013). Altricial rodents venture out of their home burrow at Post Natal Day (PND) 28, leaving the care of their adult conspecifics, and learn how to acquire nutrients and safety independently (Galef, 1981). The adolescent (PND 28–60) behavioral profile is characterized by increased risk-taking, reward- and novelty-seeking, as well as an augmented need for social and environmental stimulation (Laviola et al., 2003; Kelley et al., 2004; Marco et al., 2011) that may have evolved to promote attainment of the necessary skills for independence (Spear, 2000).

Some of the behavioral patterns common to adolescents across species may result from alterations in reward valuation marked by an increased sensitivity to reinforcers and reduced sensitivity to costs associated with obtaining them, or stimulus-reward associative learning. From a neurodevelopmental perspective, the adolescent period is characterized by pronounced changes in the functional organization and connectivity of the amygdala-prefrontal cortex (PFC) circuit (Cunningham et al., 2002, 2008) and mesocorticolimbic dopamine system (Gelbard et al., 1989; Tarazi and Baldessarini, 2000). Dopaminergic neurotransmission within the striatum and PFC is critical to adaptive reward learning and motivation (Berridge and Robinson, 1998; Salamone and Correa, 2002; Cagniard et al., 2006; Ostlund et al., 2011; Salamone et al., 2012; Richard et al., 2013). D1-like (D1R) receptor signaling contributes to cortico-striatal plasticity and regulates reward learning and effort-based decision making (Beninger and Miller, 1998; Baldwin et al.,

2002; Schweimer and Hauber, 2006). Similarly, D2-like (D2R) receptor-mediated signaling in the striatum has been linked to effort expenditure toward palatable rewards (Trifilieff et al., 2013) and learning from positive outcomes (Groman et al., 2011). The adolescent period is marked by extensive pruning of dopamine D1R and D2R in the striatum (Gelbard et al., 1989; Teicher et al., 1995; Tarazi and Baldessarini, 2000) that may be associated with behavioral differences in reward choices in adolescent vs. adult animals.

Connections between the amygdala and PFC are critical for reward responses and choices between options of different value (Baxter et al., 2000; Blair et al., 2006; Waraczynski, 2006). Structural remodeling within this circuit may be partially dependent on neural cell adhesion molecule (NCAM) function. Previous research has shown that polysialylated NCAM (PSA-NCAM) is critical in synaptic remodeling and plasticity (Muller et al., 1996; Durbec and Cremer, 2001) and modulates cortical neuron sensitivity to neurotrophins (Vutskits et al., 2001). It is expressed in brain regions undergoing structural reorganization during development and in adulthood, including the hippocampus, amygdala, and frontal cortex (Nacher et al., 2002a,b; Seki, 2002; Varea et al., 2005). Interestingly, dopamine signaling and PSA-NCAM expression show bidirectional interactions: manipulations of dopamine signaling (systemically and in the medial frontal cortex) have been linked to alterations in PSA-NCAM expression (Castillo-Gómez et al., 2008), and a role for PSA-NCAM in dopamine signaling-induced plasticity of frontocortical inhibitory circuits has also been suggested (Nacher et al., 2013). Similarly, NCAM can promote D2R internalization and subsequent degradation as well as modulate receptor-mediated signaling and behavior (Xiao et al., 2009). PSA-NCAM has already been implicated in learning and stress responses (Pham et al., 2003; Cordero et al., 2005; Bisaz et al., 2009). However, most of the work to date has focused on its role in aversive learning and fear memory, and largely centered on

hippocampal function (Senkov et al., 2006; Lopez-Fernandez et al., 2007; Kochlamazashvili et al., 2010). It is not known if PSA-NCAM also contributes to appetitive responses and if the regional specificity of its expression is developmentally specific. This molecular target is of a particular interest as NCAM polysialylation has been linked to neurodevelopmental predisposition to schizophrenia (Hildebrandt et al., 2009), abnormal social interaction and aggression (Calandreau et al., 2010), and the individual risk for alcohol-related behaviors (Barker et al., 2012). Therefore, in the present set of experiments, we directly compared adult and adolescent animals on tasks measuring both stimulus-reward associative learning and reward-cost valuation. Additionally, we examined developmental differences in dopamine D1R and D2R expression in the striatum and frontal cortex as well as PSA-NCAM expression in the frontal cortex and amygdala in adolescent and adult animals trained on an effortful reward valuation task (Stolyarova et al., 2015).

Methods

Subjects

Subjects were 24 (Adult = 12, Adolescent = 12) male Long Evans rats (Charles River Laboratories), pair housed. Adolescent animals arrived at our facility at PND 25, and adults arrived at our facility at PND 62. Adolescent animals were PND 28 and adult animals were PND 65 at the beginning of handling. Vivaria were maintained under a 12/12 h light/dark cycle at 22°C. All behavioral testing took place 5–7 days a week between 08:00 and 16:00 h during the rats' inactive period, consistent with previous and ongoing studies in our lab. Research protocols were approved by the Chancellor's Animal Research Committee at the University of California, Los Angeles.

Handling and Food Restriction

Rats were left undisturbed for 3 days after arrival to our facility to acclimate to the vivarium. Each rat was then handled for a minimum of 10 min once per day for 5 days. Animals were food-restricted to ensure motivation to work for food for a week prior and during the behavioral testing, while water was available ad libitum. All rats were food restricted based on their baseline food intake that was assessed after the animals had already acclimated to the vivarium to control for the effects of stress on feeding behavior. Food availability was gradually decreased starting with 80% of baseline intake. The amount of food given was never lower than 50% baseline. Weights were monitored daily. We ensured that adult animals did not fall below 85% of their free-feeding body weight and adolescent animals fell within normal age-matched growth weights provided by the vendor. Although there was an initial weight loss in the adult group (average maximal weight loss = 11% of baseline), both age groups showed an increase in weight by the end of the study (main effect of time: $F(18,396) = 30.843$, $p < 0.001$), which likely

resulted from the supplemental nutrition obtained from the rewards earned during testing. As expected, the average weight gain in adolescent animals was higher than in adults (t(22) = 6.82, p < 0.0001). On the two last days of food restriction prior to behavioral training, rats were fed twenty 1/2 froot loops or sugar pellets in their home cage to accustom them to the food rewards.

Stimulus-Reward Learning in Adolescent and Adult Animals

Behavioral Apparatus

Behavioral testing was conducted in operant conditioning chambers (Model 80604 Lafayette Instrument Co., Lafayette, IN) that were housed within sound- and light- attenuating cubicles. Each chamber was equipped with a house light, tone generator, video camera, and LCD touchscreen opposing the pellet dispenser. The pellet dispenser delivered 45-mg dustless precision sucrose pellets. Software (ABET II TOUCH) controlled touchscreen stimuli presentation, tone generation, tray- and house-light illumination and pellet dispensation.

Behavioral Training

Reward learning was assessed on tasks commonly used as pre-training stages for discrimination learning testing. The training protocol was adapted from Kosheleff et al. (2012) and Izquierdo et al. (2010). Due to a short duration of adolescence in rats (i.e., PND 28–60), only two of the initial phases were used in the present experiment: Initial Touch Training (ITT) and Must Touch Training (MTT). During habituation, rats were required to eat five pellets out of the pellet dispenser inside of the chambers within 15 min before exposure to any stimuli on the touchscreen. ITT began with the display of white graphic stimuli on the black background at the bottom of the touchscreen. The stimuli measured 45×45 mm^2 and were within reach for both

adult and adolescent animals. During this stage a trial could be terminated for one of two reasons: if a rat touched the displayed stimulus or if the stimulus display time (40 s) ended, after which the stimulus was removed and black background displayed. The disappearance of the image was paired with the onset of a "reinforcer event": dispensation of one (low reward, LR; at the termination of stimulus time) or three (high reward, HR; stimulus touched) sucrose pellets, a 1 s tone, and an illumination of the tray-light. Trials were separated by 10 s ITI. In MTT, a trial could be terminated only if the rat touched the image, which then disappeared followed by reward delivery (**Figure 2-1**). For both ITT and MTT, the criterion for advancement was set to 60 rewards consumed in 45 min. Animals were given one 45 min session per day until the criterion was reached.

Reward-Cost Valuation in Adolescent and Adult Animals

Behavioral Testing Apparatus

Rats were tested on a task previously described in Stolyarova et al. (2015) which utilized a maze with three possible courses of action, each associated with different effort requirements and reward magnitudes. Behavioral training and testing were conducted in a standard eight-arm radial maze with arms extending from a central arena with a diameter of 25 cm. Arms were 50 cm long and 12 cm wide. The positions of extramaze cues remained constant throughout all phases of the experiment. The four arms nearest the start arm were permanently blocked, leaving a start arm and three choice arms accessible to animals. One arm of the maze was randomly designated as a low effort/reward (LER), another as a medium effort/reward (MER), and the third as a high effort/reward (HER) arm. The arm assignment was counterbalanced across animals and held constant between sessions. The barrier heights associated with MER and HER options were

adjusted for the present experiment compared to previous study due to the reduced ability of adolescent animals to climb over the tallest 30 cm barrier. The arm containing the low reward was unimpeded by a barrier, but in order to obtain the medium or high reward, rats were required to climb a 15 cm or 25 cm barrier, respectively. Rats were required to climb straight up the side (90°) and down at an angle to the food reward located at the end of the goal arm. "froot loops" (Kellogg NA Co., Battle Creek, MI) were given as food rewards during testing: a "high reward" consisted of four 1/2 froot loops (i.e., two froot loops), a "medium reward" consisted of two 1/2 froot loops (1 froot loop), and a "low reward" consisted of one 1/2 froot loop. Between trials, the rat was removed from the maze and placed in clear Plexiglas holding chamber with a 1651 cm^2 base and 38.1 cm walls.

Habituation

A habituation and training protocol adapted from Walton et al. (2002) was used to habituate the rats to the maze and familiarize them with the froot loops. During the acclimation phase 5 1/2 froot loops were placed into each arm of the maze (20 total). Each rat was individually placed into the maze and allowed to explore and eat froot loops freely. Criterion for advancement to the next phase was consumption of 20 1/2 froot loops within 15 min.

Reward Magnitude Training. Phase 1

In this phase, one goal arm was baited with four 1/2 froot loops (HR arm), another with two 1/2 froot loops (MR arm), and the third arm with one 1/2 froot loop (LR arm). The arm assignment was counterbalanced across animals, and held constant between sessions. Rats were allowed to sample freely from all arms for ten trials. No barriers were present at this phase. Each

trial lasted until the rat finished all the froot loops. Trials were separated by a 30 s inter-trial interval (ITI), during which time they were placed in an empty holding chamber. The order of arm visits was recorded. Criterion for advancement to the next phase was completion of ten trials within 30 min.
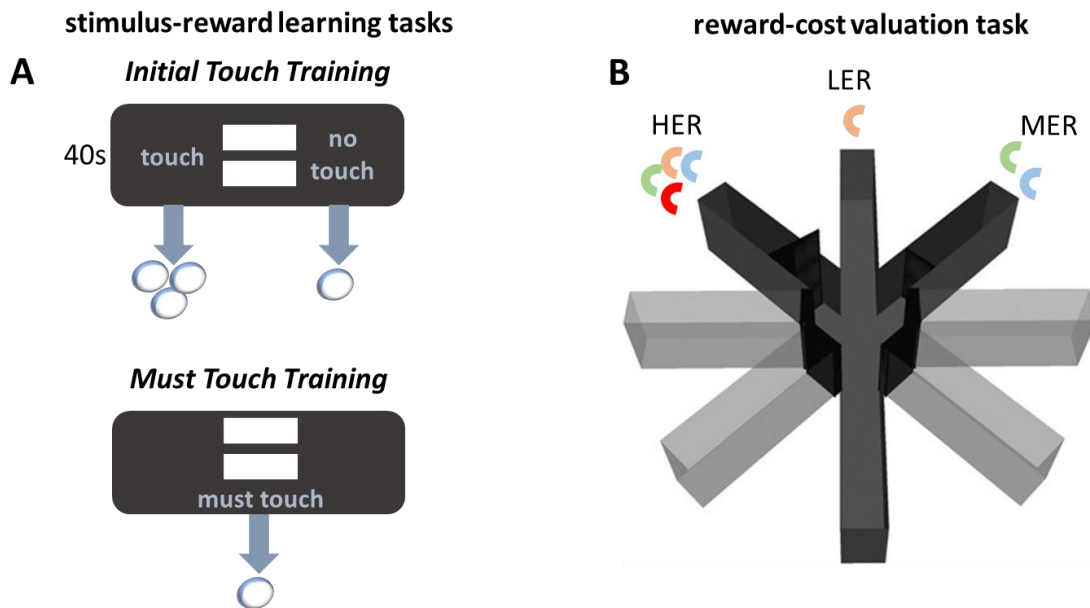
Figure 2-1. Reward valuation and learning tasks. A. Stimulus-reward associative learning tasks. In Initial Touch Training (ITT), a trial could be terminated for one of two reasons: if a rat touched the displayed stimulus, or if the stimulus display time (40 s) ended, after which the stimulus was removed and black background displayed. The disappearance of the image was paired with the onset of a "reinforcer event": dispensation of one (low reward, LR; at the termination of stimulus time) or three (high reward, HR; stimulus touched) sucrose pellets, a 1 stone, and an illumination of the tray-light. In Must Touch Training (MTT), a trial could be terminated only if the rat touched the image, which then disappeared followed by reward delivery. **B. Effortful reward valuation task.** Behavioral training and testing were conducted in a standard eight-arm radial maze, with arms extending from a central arena with a diameter of 25 cm. Arms were 50 cm long and 12 cm wide. The four arms nearest to the start arm were permanently blocked, leaving a start arm and three choice arms accessible to animals. One arm of the maze was randomly designated as a low-effort/reward (LER) arm, another as a medium-effort/reward (MER) arm, and the third as a high-effort/reward (HER) arm. The arm assignment was counterbalanced across animals, and held constant between sessions. The arm containing low reward was unimpeded by a barrier, but to obtain a medium or high reward, rats were required to climb a 15- or 25 cm barrier, respectively. 'Froot loops' were given as food rewards during testing: a 'high reward' consisted of four ½ froot loops (ie, two froot loops), a 'medium reward' consisted of two ½ froot loops (one froot loop), and a 'low reward' consisted of one ½ froot loop.

Reward Magnitude Training. Phase 2

This phase was similar to Phase 1 of reward magnitude training, except that animals were allowed to visit only one arm per trial. Rats were removed from the maze as soon as the arm was chosen and the reward was consumed. Animals were given 10 trials per day separated by a 30 s ITI. This phase marked the beginning of learning to visit only one arm as well as continuing to learn each arm's associated reward values. Criterion for advancement to the next phase was choice of HR arm on 80% or more of the trials for two consecutive days.

Alternating Free/Forced Choice Trials with Barriers

During this phase, rats were required to climb barriers to achieve higher rewards. The LER arm continued to be unimpeded by a barrier, but in order to obtain the medium (MER) or high (HER) reward, rats were required to climb a 15 cm or 25 cm barrier, respectively. Thirteen trials were administered per day. Each day of testing consisted of ten free and three forced choice (one for each arm) trials, administered at the beginning. Thus, the structure of the testing was as follows: forced choice trials (1 through 3), followed by ten free choice trials (4 through 13). On forced choice trials all goal arms except one were blocked. The order of arm presentation during forced choice trials was counterbalanced between days. Upon eating the food reward, the rat was placed in a holding chamber for a 30 s ITI, during which the maze was wiped clean with 70% ethanol to prevent the rat's use of scent-guided choice. Rats were tested daily until stable baseline choice performance was established (choice preferences on free choice trials did not differ across three consecutive days).

Amygdalar and Frontocortical PSA-NCAM and Striatal D1R and D2R Expression in Adolescent and Adult Animals

Tissue Dissection

Rats tested on a maze task were euthanized 1d after the last day of behavioral testing (late adolescent, PND 50 = 8; adult, PND 86 = 8) with an overdose of sodium pentobarbital (250 mg/kg, i.p.) and decapitated. The brains were immediately extracted and two millimeter-thick coronal sections of frontal cortex, striatum, and amygdala were further rapidly dissected, using a brain matrix, over wet ice at 4°C. Frontocortical dissections included ventral (orbital) and medial sectors of the frontal cortex, but excluded most lateral, posterior (agranular insular) regions. Striatal dissections included both dorsal and ventral subregions. Following dissection, samples were immersed in isopentane (surrounded by dry ice) and then stored at −80 °C before homogenization.

ELISA Method

To prepare the tissue for the assays, 0.3 mL (frontal cortex, striatum) or 0.2 mL (amygdala) of PBS (0.01 mol/L, pH 7.2) containing a protease and phosphatase inhibitor cocktail (aprotinin, bestatin, E-64; leupeptin, NaF, sodium orthovanadate, sodium pyrophosphate, β-glycerophosphate; Thermo Scientific, Rockford, IL) was added to each sample. Each tissue sample was minced, homogenized, sonicated with an ultrasonic cell disrupter, and centrifuged at 5,000 g at 4°C for 10 min. Supernatants were removed and stored at −20°C until ELISA assays were performed. Bradford protein assays were also performed to determine total protein concentrations in each sample. D1R, D2R (Cat# SEB299Ra and SEA673Ra, Cloud-Clone Corp., Houston, TX) and PSA-NCAM (Cat# 67-ABC0027B, ALPCO Diagnostics, Salem, NH) protein levels were determined using commercially-available ELISA kits. The assays were performed according to

the manufacturer's instructions. The sensitivity of the assays is 0.055 ng/mL for D1R, 0.112 ng/mL for D2R, and 0.25 ng/mL for PSA-NCAM, and the detection range is 0.156–10 ng/mL for D1R, 0.312–20 ng/mL for D2R, and 0.25–16 ng/mL for PSA-NCAM. The concentration of each protein is presented as ng/mg of total protein accounting for dilution factor.

Data Analyses

The data were analyzed with mixed-effects generalized linear models (GLM) in MATLAB (fitglme function; Statistics and Machine Learning Toolbox; MathWorks, Natick, Massachusetts; Versions R2017a and R2020a). ITT touch rate, effortful choice and protein expression data generally met the normality assumption. The developmental group (adolescent vs adult, categorical) was modeled as a fixed factor. For effortful choice and latency data, the trial type (LER, MER vs HER) was added as an additional fixed predictor and animal ID was included as a random factor. Sessions to criterion data across tasks were assumed to be Poisson distributed. We judged the significance of each predictor based on the t-test of the associated beta coefficient. Statistical significance was noted when p-values were less than 0.05.

Results

Stimulus-Reward Learning

In the ITT and MTT stages of training, animals learned to associate the visual stimulus presented on the touchscreen with the sucrose reward: they needed to learn to identify the relevant stimulus among other cues and objects present in the testing chamber, respond appropriately to the stimulus (i.e., with the nosepoke), and anticipate reward delivery in the magazine. Both adolescent and adult animals readily completed the ITT and MTT (**Figure 2-2**).
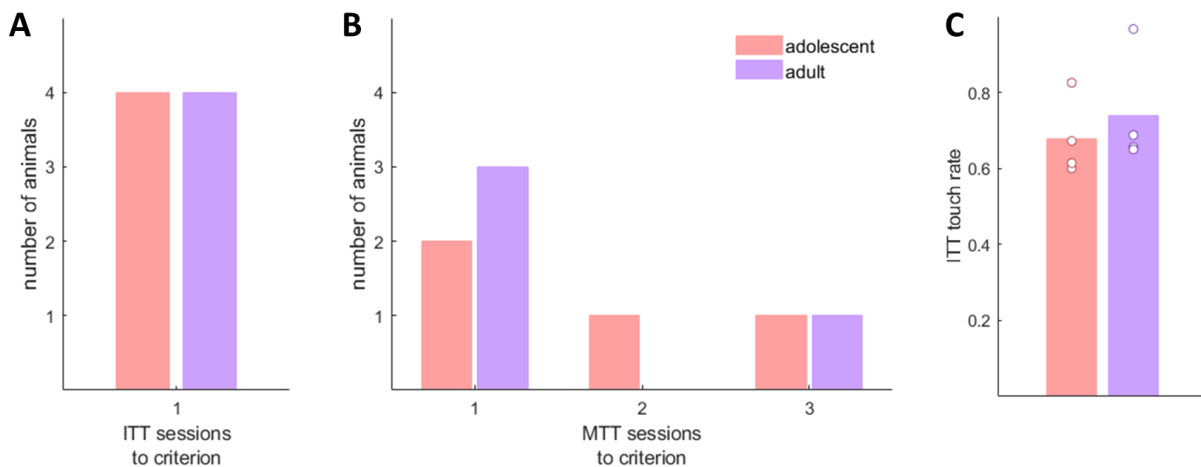
Figure 2-2. Adolescent animals acquire stimulus-reward associations as readily as adults.
A. The ITT criterion was reached in one session by animals in both groups. **B, C.** There were no statistically significant developmental group differences in the number of sessions to reach the MTT criterion or in the average touch rate on the ITT task. **A, B.** Bars represent the number of animals that required a given number of days to reach the training criterion. **C.** Bars represent group averages and dots represent individual animal data.

The ITT criterion was reached in one session by all animals in both groups, whereas the range for MTT completion was 1–3 sessions. There were no developmental group differences in sessions to criterion on MTT [β=0.15415, t(6)=0.27708, p=0.79102]. We also analyzed the touch rate on the ITT (the ratio of the number of trials terminated due to the rat touching the stimulus to the number of trials terminated at the end of the maximum stimulus duration without the rat

touching the stimulus) and found no developmental differences [β=-0.062213, t(6)=-0.78233, p=0.46377]. These results combined suggest that adolescent animals are not different from adults in learning about simple stimulus-reward contingencies.
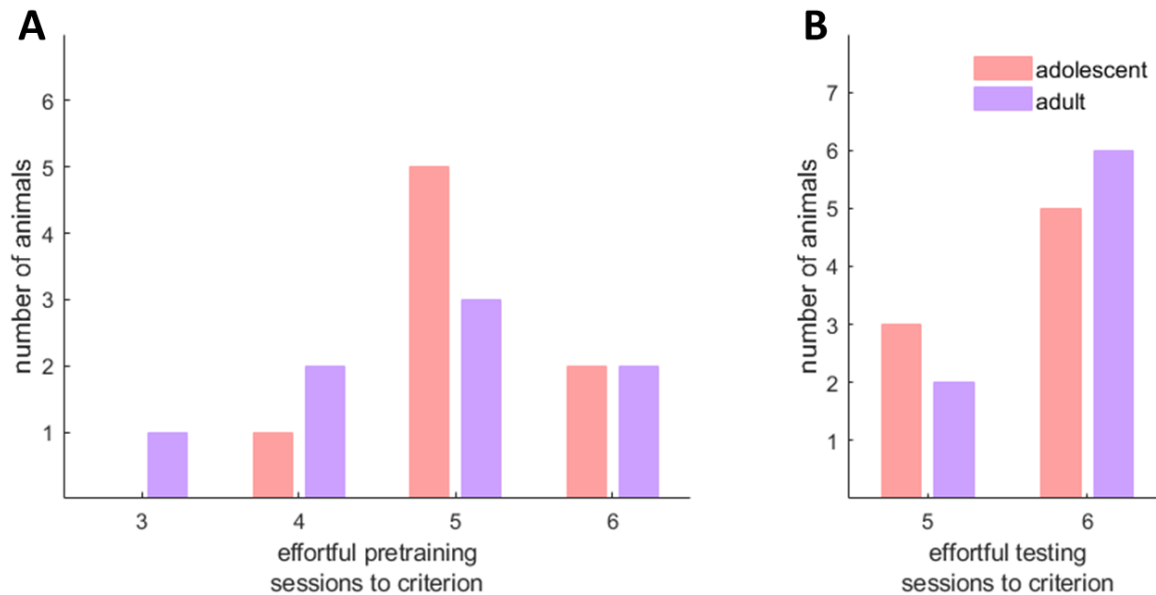
Figure 2-3. Adolescent animals learn the spatial distribution of rewards and costs on a maze task as readily as adults. A, B. There were no statistically significant developmental group differences in the number of sessions to reach either the pretraining or training criterion on the maze task. Bars represent the number of animals that required a given number of days to reach the criterion.

Reward-Cost Valuation

Congruent with our findings on the stimulus-reward learning task, adolescent animals were as swift as adults to learn the spatial distribution of rewards and effort costs in their maze environment: there were no developmental group differences in the number of sessions to reach either the pretraining [β=-0.021979, t(14)=-0.10483, p=0.918] or training [β=0.075986, t(14)=0.33745, p=0.74079] criterion on the effortful maze task (**Figure 2-3**).

Once preferences stabilized, we observed developmental differences in effortful choice behavior [**Figure 2-4**; developmental group x trial type interaction: β=0.18958, t(44)=3.0818, p=0.0035439]. Compared to adults, adolescent animals chose the LER option less [β=-0.29167, t(14)=-6.2989, p=1.959*10^{-05}] and the MER option more [β=0.2, t(14)=3.0729, p=0.0082657] frequently. There was also a statistical trend for the adolescent animals to choose the HER option more frequently [β=0.0875, t(14)=1.8681, p=0.082827]. Overall, adolescent animals were more sensitive to rewards and more likely to invest physical effort to procure the desirable outcomes.
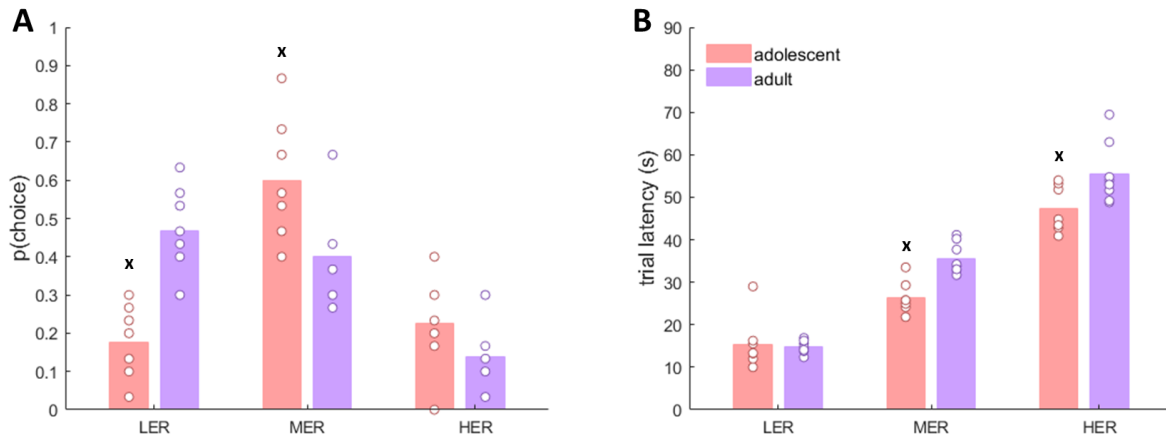
Figure 2-4. Adolescent animals invest more physical effort into obtaining rewards. A. While adult animals demonstrated a clear pattern of effort discounting, choosing options less frequently with increases in effort demands, adolescent animals were willing to invest more effort to procure larger amounts of reward. Compared to adults, adolescents chose the LER option significantly less frequently and the MER option significantly more frequently. Adolescents also chose the HER option more frequently, but this effect was only marginally significant (p=0.082827). **B.** While trial latencies increased with the barrier height in all animals, adolescent animals were significantly faster to complete MER and HER trials compared to adults. Bars represent group averages and dots represent individual animal data. x<0.05.

Trial latencies increased with barrier height in both adolescent and adult animals [**Figure 2-4**; β=20.369, t(43)=16.597, p=2.6098*10^{-20}]. Analyses of latencies also detected a significant developmental group x trial type interaction [β=-4.7868, t(43)= -2.4749, p=0.017346]: while

adolescent animals did not differ from adults in latency to complete LER trials [β=0.58321, t(14)=0.29219, p=0.77443], they took significantly less time to complete both MER [β=-9.2201, t(14)=-5.5024, p=7.7894*10^{-05}] and HER [β=-8.196, t(13)=-2.636, p=0.020551] trials.
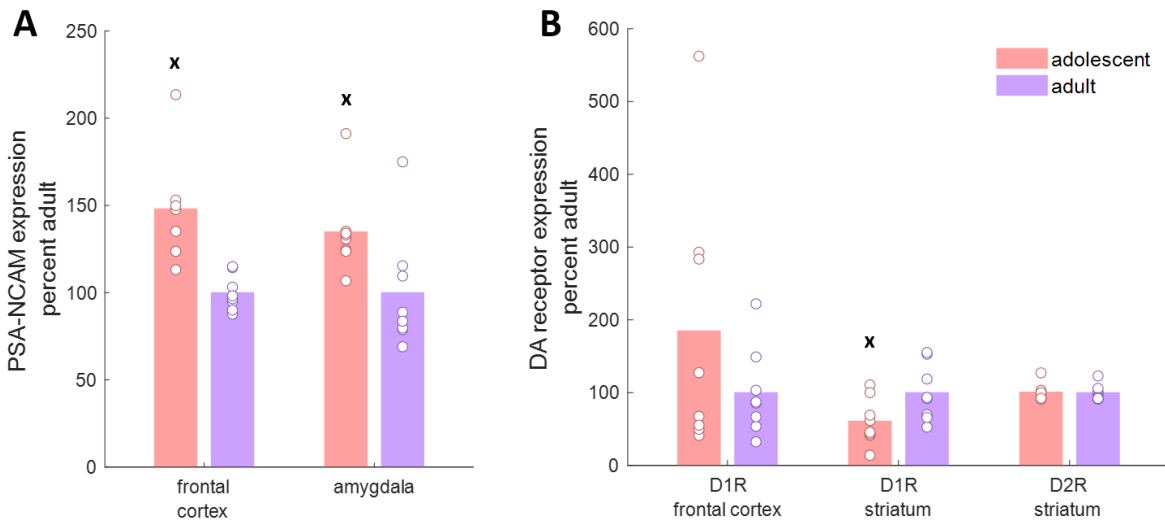
Figure 2-5. Developmental differences in PSA-NCAM, D1R and D2R expression. A. We found that PSA-NCAM was significantly upregulated both in the frontal cortex and amygdala of late adolescents compared to adult animals. **B.** Striatal D1R expression was significantly lower in adolescent compared to adult animals. However, we did not find developmental differences in frontocortical D1R or striatal D2R expression. Bars represent group averages and dots represent individual animal data. x<0.05.

PSA-NCAM, D1R and D2R expression

We quantified the expression of PSA-NCAM in the frontal cortex and amygdala, D1R in the frontal cortex and striatum, and D2R in the striatum in animals that had been previously tested on the effortful maze task. We found that PSA-NCAM was upregulated both in the cortex [β=48.032, t(13)=4.2895, p=0.00088032] and amygdala [β=34.935, t(14)=2.5119, p=0.024886] of

late adolescent compared to adult animals (**Figure 2-5**). The D1R expression was reduced in the striatum of late adolescents compared to adults [β=-39.298, t(14)=-2.356, p=0.033575], but there were no developmental differences in frontocortical D1R expression [β=84.906, t(14)=1.3269, p=0.20578]. Our analyses of the striatal D2R expression revealed no developmental group differences [β=0.78109, t(14)=0.15064, p=0.88241].
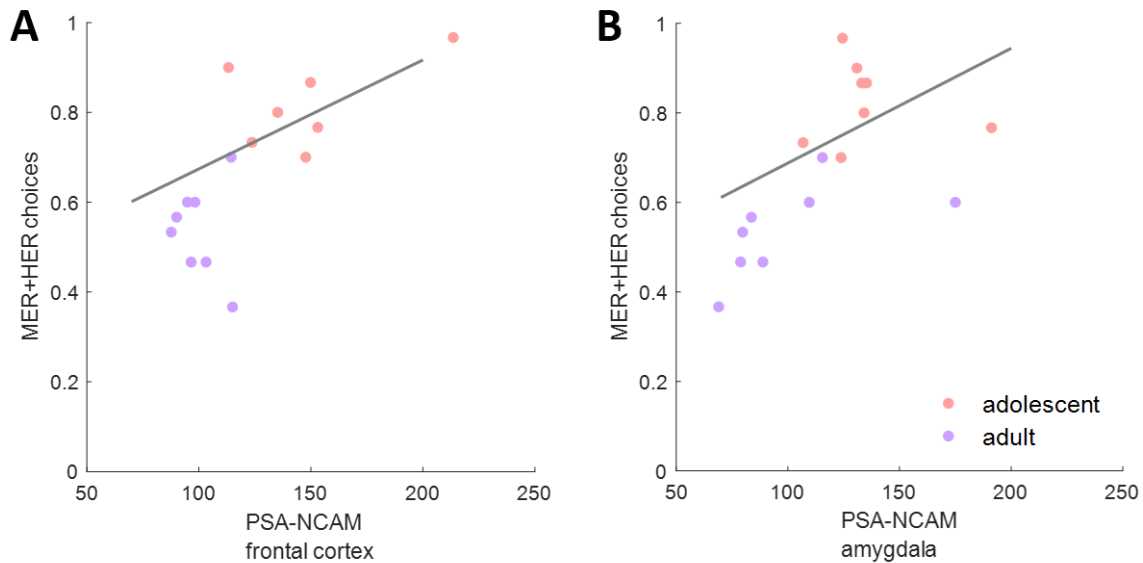
Figure 2-6. PSA-NCAM in the frontal cortex and amygdala predicts effortful choices. A, B. PSA-NCAM both in the frontal cortex and the amygdala was significantly associated with the effortful choice (MER and HER options combined) probability, over and above the (non-significant) effects of other predictors (including the frontocortical D1R, striatal D1R, and striatal D2R). Solid regression lines represent the statistically significant relationship between PSA-NCAM and the choice behavior, over and above the effects of other predictors. Because coefficients used in the regression equation were based on GLM analyses that included all neural predictors, the regression lines reflect the association between PSA-NCAM and effortful choice probability after accounting for the effects of other predictors. Dots represent individual animal data.

PSA-NCAM in the Cortex and Amygdala Predicts Effortful Choices

Finally, we conducted a GLM analysis to identify which of the proteins assessed in the present study were more reliably associated with effortful choice performance. The criterion/dependent variable in this analysis was the percent choice of the MER and HER, both effortful, options combined. The measures of cortical PSA-NCAM, amygdalar PSA-NCAM, frontocortical D1R, striatal D1R and striatal D2R were all entered as fixed continuous predictors. Among these predictors, only PSA-NCAM, both in the frontal cortex and the amygdala, was significantly associated with choice behavior, over and above the (non-significant) effects of other predictors (**Figure 2-6**). PSA-NCAM within the frontal cortex [β=0.0024261, t(9)=2.5373, p=0.031852] and amygdala [β=0.0025633, t(9)=2.9225, p=0.016961] was positively associated with the percent choice of effortful options (MER and HER combined); no other neural predictors showed a significant relationship with MER+HER probability [cortical D1R: β=0.00016229, t(9)=0.51267, p=0.62053; striatal D1R: β=-0.00028328, t(9)=-0.35013, p=0.73429; striatal D2R: β=-0.0035417, t(9)=-1.2132, p=0.25592].

Discussion

Optimal behavioral strategies may have a distinct meaning for adolescent and adult animals, depending on the temporal proximity of the reproductive period (Gluckman et al., 2007). For example, increased energy expenditure toward palatable foods may be suboptimal in adult animals that need to invest more time and effort into searching for potential mates or providing care and shelter for offspring. Conversely, foraging for nutritional rewards is critical for adolescent animals to ensure immediate survival, promote growth and increase reproductive fitness later in life. The present findings inform our understanding of behavioral phenotypes at different developmental stages. Specifically, we show that adolescent animals are indistinguishable from adults in a simple form of appetitive associative learning but exhibit an enhanced motivation to invest effort to obtain larger rewards over less valuable, freely available options. Additionally, we report different expression patterns of frontocortical and amygdalar PSA-NCAM and striatal dopamine receptors depending on developmental period.

Reward Learning is Similar in Adolescent and Adult Animals

Associative learning is highly important for many characteristic animal behaviors in the wild, including exploration of novelty, increased attention to change, and approach to potential rewards (Cloninger and Gilligan, 1987). An ability to learn the association between appetitive outcomes and predictive stimuli provides an evolutionary advantage as it allows animals to maximize rewards of great importance to mammalian species (Bitterman, 1975). The results of the present investigation suggest that by the time rodents transition from complete reliance on their adult conspecifics and begin exploring their surroundings independently (i.e., adolescent period; Galef, 1981), they already possess associative reward learning skills. We observed no age

differences in basic forms of stimulus-reward and instrumental learning: adolescent animals were as fast as adults to master the operant task and direct their responses toward the relevant stimuli. Similarly, they efficiently learned the spatial distribution of reward densities in the maze task and established stable choice behavior at a rate comparable to adults.

Although we did not examine reward learning earlier in development, the present data suggest that appetitive learning is established before adolescence. It needs to be noted, however, that one previous report demonstrated impaired odor-discrimination learning in adolescent compared to both juvenile and adult animals (Garske et al., 2013). In that task, which may be more ethologically relevant for rodents than our visual task, animals were first trained to dig in a cup filled with unscented playground sand to obtain a palatable food reward, after which they were presented with two odorized cups only one of which contained the reward. Adolescent animals were slower to acquire this odor-association task, an effect that disappeared with pre-training during the juvenile period. Taken together, these results suggest that adolescent animals are not different on measures of simple appetitive reward learning; they were still able to acquire sand-reward association. However, rats displayed a limited ability to fine-tune cue representations and demonstrated learning difficulties when cues had more than one attribute. Previous reports also indicate compromised ability to behaviorally adapt to a change in operant contingencies and extinguish previously reinforced responding in adolescent animals (Sturman et al., 2010; Andrzejewski et al., 2011). However, similar to the present results, younger animals in both of these studies efficiently learned simple stimulus- and action-reward associations.

Adolescents Invest More Effort into More Profitable Options

Adolescent animals displayed increased motivation to work for rewards of greater magnitude. Their choice behavior was characterized by increased effort expenditure toward larger rewards, while adult animals showed a clear pattern of reward devaluation as a function of increased barrier heights. The observed differences in choice preferences may be due to potentiated reactions to novel palatable foods in younger animals. Adolescents have been previously shown to display conditioned place preference, a measure of reward, to novelty, an effect that is absent in adults (Douglas et al., 2003). Additionally, adolescent animals are more sensitive to natural rewards, consume more sucrose solution and exhibit greater positive taste responses than their adult counterparts (Wilmouth and Spear, 2009). Alternatively, adolescents may be more sensitive to changes in unpredictable conditions in their habitat, which modulates effort expenditure (McNamara et al., 2013). Specifically, in the present study, both adult and adolescent animals were raised in a benign, nutritionally optimal environment, with food and water provided ad libitum, and were socially housed; they did not need to actively forage for rewards. The mismatch that was introduced as a result of the short-term food restriction may have had a more profound impact on adolescents compared to adults. These findings suggest that experiences during adolescence may have more potential adaptive significance than those encountered later in adulthood.

Neurodevelopmental Correlates of Reward-Sensitive Endophenotype

Dopaminergic neurotransmission within the striatum has long been recognized as critical for incentive motivation and optimal response allocation to rewards (Berridge and Robinson, 1998; Salamone and Correa, 2002; Ostlund et al., 2011; Salamone et al., 2012; Richard et al., 2013). D1R and D2R density in the striatum peaks at the onset of the adolescent period, followed by

extensive pruning to adult levels (Gelbard et al., 1989; Teicher et al., 1995; Tarazi and Baldessarini, 2000). The results of the present investigation revealed unaltered D2R expression but reduced D1R expression in the striata of animals that had experiences with reward learning during adolescence compared to animals that went through the same learning in adulthood. Decreased expression of D1R may result in diminished neuronal excitability in the striatonigral pathway upon dopamine release (Aosaki et al., 1998), and may ultimately lead to reduced learning from positive outcomes (Cox et al., 2015). It needs to be noted that because brains were collected following training and establishment of stable performance, we are unable to distinguish age-specific from experience-dependent receptor expression profiles. However, previous reports indicate that D1R expression reaches mature levels by early adulthood (Teicher et al., 1995; Tarazi and Baldessarini, 2000). Therefore, reward experiences during adolescence may exaggerate normal pruning patterns and result in lower D1R levels as compared to the same experiences encountered in adulthood. Increased D1R expression early in adolescence (Gelbard et al., 1989; Tarazi and Baldessarini, 2000) may aid in establishing a pattern of behavior characterized by greater effort expenditure toward larger rewards, whereas decreased levels of D1R expression at the onset of adulthood would render animals less vulnerable to the effects of experiences with potent reinforcers.

Information transfer between the amygdala and PFC has been shown to be critical for optimal reward-driven effort expenditure in maze tasks (Floresco and Ghods-Sharifi, 2007), with basolateral amygdala (BLA) signaling differences in reward magnitude (Salinas et al., 1993; Pratt and Mizumori, 1998). Projections from the BLA to the frontal cortex undergo remarkable development during adolescence (Casey et al., 2000; Cunningham et al., 2002, 2008; Brenhouse and Andersen, 2011). PSA-NCAM may play an important role in such structural and functional

changes given its importance in activity-dependent synaptic remodeling and developmental events (Muller et al., 1996, 2010; Dey et al., 1999; Durbec and Cremer, 2001; Kiss and Muller, 2001; Welzl and Stork, 2003), resulting in prominent patterns of expression in regions undergoing active functional restructuring (Nacher et al., 2002a,b; Seki, 2002; Varea et al., 2005). Tsoory et al. (2008) reported significant decreases in PSA-NCAM expression with developmental progression from adolescence into adulthood in the amygdala and hippocampus of naïve animals. The results of the present investigation revealed increased levels of PSA-NCAM expression in the frontal cortex and amygdala in late adolescent compared to adult animals that had been trained on an effortful reward valuation task. Furthermore, we found that PSA-NCAM levels in the frontal cortex and amygdala are predictive of a pattern of reward-directed effort expenditure, consistent with a reward-sensitive endophenotype. To our knowledge, this is the first report showing a link between outcome valuation and developmentally-specific differences in PSA-NCAM expression. PSA-NCAM in the adult brain is restricted to interneurons, at least in the frontal cortex and BLA, and may aid in the incorporation of interneurons into the circuitry to modulate local inhibition (Gascon et al., 2007; Gómez-Climent et al., 2011; Nacher et al., 2013). Increased levels of PSA-NCAM expression in adolescent animals in the present study may index increased structural plasticity within these brain regions and represent a neural correlate of a reward-sensitive endophenotype. However, additional investigations utilizing direct manipulations targeted to adolescent BLA and subregions within the frontal cortex are needed to establish a causal role for PSA-NCAM in adolescent-specific behavioral traits.

References

Andrzejewski, M. E., Schochet, T. L., Feit, E. C., Harris, R., McKee, B. L., and Kelley, A. E. (2011). A comparison of adult and adolescent rat behavior in operant learning, extinction and behavioral inhibition paradigms. *Behav Neurosci* 125, 93–105.

Aosaki, T., Kiuchi, K., and Kawaguchi, Y. (1998). Dopamine D1-like receptor activation excites rat striatal large aspiny neurons *in vitro*. *J Neurosci* 18, 5180–5190.

Baldwin, A. E., Sadeghian, K., and Kelley, A. E. (2002). Appetitive instrumental learning requires coincident activation of NMDA and dopamine D1 receptors within the medial prefrontal cortex. *J Neurosci* 22, 1063–1071.

Barker, J. M., Torregrossa, M. M., and Taylor, J. R. (2012). Low prefrontal PSA-NCAM confers risk for alcoholism-related behavior. *Nat Neurosci* 15, 1356–1358.

Baxter, M. G., Parker, A., Lindner, C. C., Izquierdo, A. D., and Murray, E. A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *J Neurosci* 20, 4311–4319.

Beninger, R. J., and Miller, R. (1998). Dopamine D1-like receptors and reward-related incentive learning. *Neurosci Biobehav Rev* 22, 335–345.

Berridge, K. C., and Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Rev* 28, 309–369.

Bisaz, R., Conboy, L., and Sandi, C. (2009). Learning under stress: a role for the neural cell adhesion molecule NCAM. *Neurobiol Learn Mem* 91, 333–342.

Bitterman, M. E. (1975). The comparative analysis of learning. *Science* 188, 699–709.

Blair, K., Marsh, A. A., Morton, J., Vythilingam, M., Jones, M., Mondillo, K., et al. (2006). Choosing the lesser of two evils, the better of two goods: specifying the roles of ventromedial prefrontal cortex and dorsal anterior cingulate in object choice. *J Neurosci* 26, 11379–11386.

Brenhouse, H. C., and Andersen, S. L. (2011). Developmental trajectories during adolescence in males and females: a cross-species understanding of underlying brain changes. *Neurosci Biobehav Rev* 35, 1687–1703.

Cagniard, B., Beeler, J. A., Britt, J. P., McGehee, D. S., Marinelli, M., and Zhuang, X. (2006). Dopamine scales performance in the absence of new learning. *Neuron* 51, 541–547. d

Calandreau, L., Márquez, C., Bisaz, R., Fantin, M., and Sandi, C. (2010). Differential impact of polysialyltransferase ST8SiaII and ST8SiaIV knockout on social interaction and aggression. *Genes Brain Behavior* 9, 958–967.

Casey, B. J., Giedd, J. N., and Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol Psychol* 54, 241–257.

Castillo-Gómez, E., Gómez-Climent, M. A., Varea, E., Guirado, R., Blasco-Ibáñez, J. M., Crespo, C., et al. (2008). Dopamine acting through D2 receptors modulates the expression of PSA-NCAM, a molecule related to neuronal structural plasticity, in the medial prefrontal cortex of adult rats. *Exp Neurol* 214, 97–111.

Cloninger, C. R., and Gilligan, S. B. (1987). Neurogenetic mechanisms of learning: a phylogenetic perspective. *J Psychiatr Res* 21, 457–472.

Cordero, M. I., Rodríguez, J. J., Davies, H. A., Peddie, C. J., Sandi, C., and Stewart, M. G. (2005). Chronic restraint stress down-regulates amygdaloid expression of polysialylated neural cell adhesion molecule. *Neuroscience* 133, 903–910.

Cox, S. M., Frank, M. J., Larcher, K., Fellows, L. K., Clark, C. A., Leyton, M., et al. (2015). Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *Neuroimage* 109, 95–101.

Cunningham, M. G., Bhattacharyya, S., and Benes, F. M. (2002). Amygdalo-cortical sprouting continues into early adulthood: implications for the development of normal and abnormal function during adolescence. *J Comp Neurol* 453, 116–130.

Cunningham, M. G., Bhattacharyya, S., and Benes, F. M. (2008). Increasing interaction of amygdalar afferents with GABAergic interneurons between birth and adulthood. *Cereb Cortex* 18, 1529–1535.

Dey, P. M., Gochfeld, M., and Reuhl, K. R. (1999). Developmental methylmercury administration alters cerebellar PSA-NCAM expression and Golgi sialyltransferase activity. *Brain Res* 845, 139–151.

Douglas, L. A., Varlinskaya, E. I., and Spear, L. P. (2003). Novel-object place conditioning in adolescent and adult male and female rats: effects of social isolation. *Physiol Behav* 80, 317–325.

Durbec, P., and Cremer, H. (2001). Revisiting the function of PSA-NCAM in the nervous system. *Mol Neurobiol* 24, 53–64.

Floresco, S. B., and Ghods-Sharifi, S. (2007). Amygdala-prefrontal cortical circuitry regulates effort-based decision making. *Cereb Cortex* 17, 251–260.

Galef, B. G. Jr. (1981). "The ecology of weaning: parasitism and the achievement of independence by altricial mammals," in *Parental Care in Mammals*, eds D. J. Gubernick and P. H. Klopfer (New York, NY: Plenum Press), 211–241.

Garske, A. K., Lawyer, C. R., Peterson, B. M., and Illig, K. R. (2013). Adolescent changes in dopamine D1 receptor expression in orbitofrontal cortex and piriform cortex accompany an associative learning deficit. *PLoS One* 8:e56191.

Gascon, E., Vutskits, L., and Kiss, J. Z. (2007). Polysialic acid-neural cell adhesion molecule in brain plasticity: from synapses to integration of new neurons. *Brain Res Rev* 56, 101–118.

Gelbard, H. A., Teicher, M. H., Faedda, G., and Baldessarini, R. J. (1989). Postnatal development of dopamine D1 and D2 receptor sites in rat striatum. *Brain Res Dev Brain Res* 49, 123–130.

Gluckman, P. D., Hanson, M. A., and Beedle, A. S. (2007). Early life events and their consequences for later disease: a life history and evolutionary perspective. *Am J Hum Biol* 19, 1–19.

Gómez-Climent, M. Á., Guirado, R., Castillo-Gómez, E., Varea, E., Gutierrez-Mecinas, M., Gilabert-Juan, J., et al. (2011). The polysialylated form of the neural cell adhesion molecule (PSA-NCAM) is expressed in a subpopulation of mature cortical interneurons characterized by reduced structural features and connectivity. *Cereb Cortex* 21, 1028–1041.

Groman, S. M., Lee, B., London, E. D., Mandelkern, M. A., James, A. S., Feiler, K., et al. (2011). Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. *J Neurosci* 31, 7291–7299.

Hildebrandt, H., Mühlenhoff, M., Oltmann-Norden, I., Röckle, I., Burkhardt, H., Weinhold, B., et al. (2009). Imbalance of neural cell adhesion molecule and polysialyltransferase alleles causes defective brain connectivity. *Brain* 132, 2831–2838.

Izquierdo, A., Belcher, A. M., Scott, L., Cazares, V. A., Chen, J., O'Dell, S. J., et al. (2010). Reversal-specific learning impairments after a binge regimen of methamphetamine in rats: possible involvement of striatal dopamine. *Neuropsychopharmacology* 35, 505–514.

Kelley, A. E., Schochet, T., and Landry, C. F. (2004). Risk taking and novelty seeking in adolescence: introduction to part I. *Ann N Y Acad Sci* 1021, 27–32.

Kiss, J. Z., and Muller, D. (2001). Contribution of the neural cell adhesion molecule to neuronal and synaptic plasticity. *Rev Neurosci* 12, 297–310.

Kochlamazashvili, G., Senkov, O., Grebenyuk, S., Robinson, C., Xiao, M. F., Stummeyer, K., et al. (2010). Neural cell adhesion molecule-associated polysialic acid regulates synaptic plasticity and learning by restraining the signaling through GluN2B-containing NMDA receptors. *J Neurosci* 30, 4171–4183.

Kosheleff, A. R., Rodriguez, D., O'Dell, S. J., Marshall, J. F., and Izquierdo, A. (2012). Comparison of single-dose and extended methamphetamine administration on reversal learning in rats. *Psychopharmacology (Berl)* 224, 459–467.

Laviola, G., Macrì, S., Morley-Fletcher, S., and Adriani, W. (2003). Risk-taking behavior in adolescent mice: psychobiological determinants and early epigenetic influence. *Neurosci. Biobehav Rev* 27, 19–31.

Lopez-Fernandez, M. A., Montaron, M. F., Varea, E., Rougon, G., Venero, C., Abrous, D. N., et al. (2007). Upregulation of polysialylated neural cell adhesion molecule in the dorsal hippocampus after contextual fear conditioning is involved in long-term memory formation. *J Neurosci* 27, 4552–4561.

Marco, E. M., Adriani, W., Ruocco, L. A., Canese, R., Sadile, A. G., and Laviola, G. (2011). Neurobehavioral adaptations to methylphenidate: the issue of early adolescent exposure. *Neurosci. Biobehav Rev* 35, 1722–1739.

McNamara, J. M., Fawcett, T. W., and Houston, A. I. (2013). An adaptive response to uncertainty generates positive and negative contrast effects. *Science* 340, 1084–1086.

Muller, D., Mendez, P., Deroo, M., Klauser, P., Steen, S., and Poglia, L. (2010). Role of NCAM in spine dynamics and synaptogenesis. *Adv Exp Med Biol* 663, 245–256.

Muller, D., Wang, C., Skibo, G., Toni, N., Cremer, H., Calaora, V., et al. (1996). PSA-NCAM is required for activity-induced synaptic plasticity. *Neuron* 17, 413–422.

Nacher, J., Blasco-Ibáñez, J. M., and McEwen, B. S. (2002a). Non-granule PSA-NCAM immunoreactive neurons in the rat hippocampus. *Brain Res* 930, 1–11.

Nacher, J., Guirado, R., and Castillo-Gómez, E. (2013). Structural plasticity of interneurons in the adult brain: role of PSA-NCAM and implications for psychiatric disorders. *Neurochem Res* 38, 1122–1133.

Nacher, J., Lanuza, E., and McEwen, B. S. (2002b). Distribution of PSA-NCAM expression in the amygdala of the adult rat. *Neuroscience* 113, 479–484.

Ostlund, S. B., Wassum, K. M., Murphy, N. P., Balleine, B. W., and Maidment, N. T. (2011). Extracellular dopamine levels in striatal subregions track shifts in motivation and response cost during instrumental conditioning. *J Neurosci* 31, 200–207.

Pham, K., Nacher, J., Hof, P. R., and McEwen, B. S. (2003). Repeated restraint stress suppresses neurogenesis and induces biphasic PSA-NCAM expression in the adult rat dentate gyrus. *Eur J Neurosci* 17, 879–886.

Pratt, W. E., and Mizumori, S. J. (1998). Characteristics of basolateral amygdala neuronal firing on a spatial memory task involving differential reward. *Behav Neurosci* 112, 554–570.

Richard, J. M., Castro, D. C., Difeliceantonio, A. G., Robinson, M. J., and Berridge, K. C. (2013). Mapping brain circuits of reward and motivation: in the footsteps of Ann Kelley. *Neurosci Biobehav Rev* 37(9 Pt. A), 1919–1931.

Salamone, J. D., and Correa, M. (2002). Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behav Brain Res* 137, 3–25.

Salamone, J. D., Correa, M., Nunes, E. J., Randall, P. A., and Pardo, M. (2012). The behavioral pharmacology of effort-related choice behavior: dopamine, adenosine and beyond. *J Exp Anal Behav* 97, 125–146.

Salinas, J. A., Packard, M. G., and McGaugh, J. L. (1993). Amygdala modulates memory for changes in reward magnitude: reversible post-training inactivation with lidocaine attenuates the response to a reduction in reward. *Behav Brain Res* 59, 153–159.

Schweimer, J., and Hauber, W. (2006). Dopamine D1 receptors in the anterior cingulate cortex regulate effort-based decision making. *Learn Mem* 13, 777–782.

Seki, T. (2002). Expression patterns of immature neuronal markers PSA-NCAM, CRMP-4 and NeuroD in the hippocampus of young adult and aged rodents. *J Neurosci Res* 70, 327–334.

Senkov, O., Sun, M., Weinhold, B., Gerardy-Schahn, R., Schachner, M., and Dityatev, A. (2006). Polysialylated neural cell adhesion molecule is involved in induction of long-term potentiation and memory acquisition and consolidation in a fear-conditioning paradigm. *J Neurosci* 26, 10888–109898.

Spear, L. (2000). The adolescent brain and age-related behavioral manifestations. *Neurosci Biobehav Rev* 24, 417–463.

Stolyarova, A., Thompson, A. B., Barrientos, R. M., and Izquierdo, A. (2015). Reductions in frontocortical cytokine levels are associated with long-lasting alterations in reward valuation after methamphetamine. *Neuropsychopharmacology* 40, 1234–1242.

Sturman, D. A., Mandell, D. R., and Moghaddam, B. (2010). Adolescents exhibit behavioral differences from adults during instrumental learning and extinction. *Behav Neurosci* 124, 16–25.

Tarazi, F. I., and Baldessarini, R. J. (2000). Comparative postnatal development of dopamine D(1), D(2) and D(4) receptors in rat forebrain. *Int J Dev Neurosci* 18, 29–37.

Teicher, M. H., Andersen, S. L., and Hostetter, J. C. Jr. (1995). Evidence for dopamine receptor pruning between adolescence and adulthood in striatum but not nucleus accumbens. *Brain Res Dev Brain Res* 89, 167–172.

Trifilieff, P., Feng, B., Urizar, E., Winiger, V., Ward, R. D., Taylor, K. M., et al. (2013). Increasing dopamine D2 receptor expression in the adult nucleus accumbens enhances motivation. *Mol Psychiatry* 18, 1025–1033.

Tsoory, M., Guterman, A., and Richter-Levin, G. (2008). Exposure to stressors during juvenility disrupts development-related alterations in the PSA-NCAM to NCAM expression ratio: potential relevance for mood and anxiety disorders. *Neuropsychopharmacology* 33, 378–393.

Varea, E., Nácher, J., Blasco-Ibáñez, J. M., Gómez-Climent, M. A., Castillo-Gómez, E., Crespo, C., et al. (2005). PSA-NCAM expression in the rat medial prefrontal cortex. *Neuroscience* 136, 435–443.

Vutskits, L., Djebbara-Hannas, Z., Zhang, H., Paccaud, J. P., Durbec, P., Rougon, G., et al. (2001). PSA-NCAM modulates BDNF-dependent survival and differentiation of cortical neurons. *Eur J Neurosci* 13, 1391–1402.

Walton, M. E., Bannerman, D. M., and Rushworth, M. F. (2002). The role of rat medial frontal cortex in effort-based decision making. *J Neurosci* 22, 10996–11003.

Waraczynski, M. A. (2006). The central extended amygdala network as a proposed circuit underlying reward valuation. *Neurosci Biobehav Rev* 30, 472–496.

Welzl, H., and Stork, O. (2003). Cell adhesion molecules: key players in memory consolidation? *News Physiol Sci* 18, 147–150.

Wilmouth, C. E., and Spear, L. P. (2009). Hedonic sensitivity in adolescent and adult rats: taste reactivity and voluntary sucrose consumption. *Pharmacol Biochem Behav* 92, 566–573.

Xiao, M. F., Xu, J. C., Tereshchenko, Y., Novak, D., Schachner, M., and Kleene, R. (2009). Neural cell adhesion molecule modulates dopaminergic signaling and behavior by regulating dopamine D2 receptor internalization. *J Neurosci* 29, 14752–14763.

Chapter 3: Learning flexibility following adolescent drug exposure

Abstract

Corticostriatal circuitry and its dopamine inputs support flexible reward learning. It is still poorly understood how prescription drug exposure in adolescence that affects the developing corticostriatal circuitry may impact reward-guided learning and cognitive flexibility in the long-term. We studied the effects of adolescent methylphenidate (MPH) and fluoxetine (FLX) exposure on discrimination and reversal learning in adulthood. Male and female rats were administered MPH, FLX, or saline beginning on postnatal day (PND) 37. An additional comparison group of animals received methamphetamine (mAMPH), an illicit drug, treatment. After the termination of treatment and following a washout period, these rats were introduced to touchscreen tasks measuring reward learning and cognitive flexibility. Following testing, we examined dopamine D1 receptor and dopamine D2 receptor expression in the striatum and polysialylated neural cell adhesion molecule (PSA-NCAM) expression in the frontal cortex and amygdala. Adolescent pretreatment with mAMPH facilitated and with MPH attenuated discrimination learning performance in males, but not females, in adulthood. During reversal, the rate of learning was reduced in males, but not females, that had been exposed to MPH and FLX in adolescence. We also found that among control animals, females were slower to master both discrimination and reversal compared to males. Lastly, our analyses of protein expression revealed an upregulation of striatal D1 receptors in adulthood following adolescent mAMPH and MPH exposure, an upregulation of striatal D2 receptors following adolescent MPH exposure and higher levels of frontocortical PSA-NCAM expression in male, compared to female, animals. These results show the enduring effects of adolescent drug exposure on learning, cognitive flexibility, and neural development in rats.

Introduction

The adolescent period is characterized by increased risk-taking, reward-seeking, and an enhanced need for environmental stimulation and exploration (Kelley et al., 2004; Laviola et al., 2003; Marco et al., 2011); characteristics that likely evolved to promote skills for independence (Spear, 2000). Changes in mesocorticolimbic dopamine (DA) signaling provide much of the basis for this behavioral phenotype. DA D1 and D2 receptor densities in the striatum peak at the onset of the rat adolescent period (postnatal day, PND 28) but decrease with maturity (Tarazi and Baldessarini, 2000; Gelbard et al., 1989; Teicher et al., 1995). We recently reported reduced D1 expression, and unaltered D2 expression, in the striata of animals that had experiences with reward learning during adolescence when compared to animals that went through the same learning in adulthood (Stolyarova and Izquierdo, 2015). This suggests that learning and the experience of cognitive training may interact with neural maturation processes to shape long-lasting expression profiles of D1 receptors in particular (Wass et al., 2013). Exposure to psychostimulants may also cause robust changes in DA receptors in the developing brain that manifest in long-lasting effects on learning and behavior in adulthood. Adolescent rats treated for 2 months with ADHD medication methylphenidate (MPH, 1 and 2 mg/kg) beginning on PND 30 show significantly reduced D2 receptor binding compared to vehicle-treated rats, as measured by microPET (Thanos et al., 2007). This is likely meaningful to behavior since low striatal D2 receptor availability has been associated with poor reversal learning and high addiction vulnerability (Izquierdo and Jentsch, 2012).

The administration of prescription drugs to adolescents is at an all-time high (Zito et al., 2000; Miech et al., 2015). Some of the most commonly-prescribed are MPH for ADHD (Shanks et al., 2015; Caprioli et al., 2015; Crawford et al., 2011; Gray et al., 2007), and fluoxetine (FLX)

for the treatment of Major Depression (Iñiguez et al., 2010; Iñiguez et al., 2014; Homberg et al., 2011). MPH exhibits a similar pharmacological profile to amphetamines and cocaine, may modulate neurodevelopment (Grund et al., 2007; Thanos et al., 2007; Adriani et al., 2006), and by extension, may impact learning and behavior mediated by the corticostriatal circuitry. There is evidence for long-term effects of adolescent MPH exposure on adult locomotor behavior and addiction vulnerability. These effects include increased sensitization to later methamphetamine administration (mAMPH; Shanks et al., 2015), increased cocaine abuse risk (Jordan et al., 2014), increased alcohol intake (Gill et al., 2014), and increased cocaine-induced reward and behavioral sensitization (Achat-Mendes et al., 2003) in adulthood (cf. Gray et al., 2007). The effects of adolescent FLX, conversely, appear limited to significant increases in anxious responding to emotion-eliciting stimuli (Iñiguez et al., 2010; Iñiguez et al., 2014; Homberg et al., 2011; cf. Norcross et al., 2008). There is relatively little known about the long-lasting effects of adolescent FLX exposure on later adult learning and behavior, though in the adult, FLX results in fewer errors in the early phase of reversal learning (Brigman et al., 2010).

To our knowledge there has not yet been a systematic comparison of the long-term effects of adolescent MPH and FLX exposure on learning flexibility and associated D1 and D2 receptor expression in the striatum in adulthood. Additionally, most studies addressing the effects of MPH and FLX overall were done on male rats. Therefore, in the present experiments we investigated the effects of adolescent MPH and FLX exposure on learning and cognitive flexibility in adulthood in both male and female rats. We also compared the effects of these prescription medications to those of escalating doses of the illicit drug mAMPH and studied the long-term effects of adolescent drug exposure on D1 and D2 receptor expression. Finally, we quantified the expression of the polysialylated neural cell adhesion molecule (PSA-NCAM) in the frontal cortex and amygdala.

While many previous studies have focused on interactions between the DA system and psychostimulant exposure in adulthood, PSA-NCAM represents a novel target. We have previously observed increased levels of PSA-NCAM expression in both the frontal cortex and amygdala of late adolescents compared to adults that were trained on an effortful reward valuation task, which may index increased structural plasticity and represent a neural correlate of a reward sensitive endophenotype (**Chapter 2**). Furthermore, PSA-NCAM and DA systems have been shown to interact bidirectionally (Castillo-Gómez et al., 2008; Nacher et al., 2013; Xiao et al., 2009).

Methods

Subjects

Male (n=16) and female (n=16) Long–Evans rats (Charles River Laboratories, Inc.) arrived on post-natal day (PND) 28, early rat adolescence (Spear 2000), weighing between 76 and 100g, and were socially housed in same-sex pairs with both males and females housed in the same room. All rats were habituated to the vivarium from PND 28 to 31. The rats were maintained on a restricted diet during behavioral testing (no less than 85% of free-feeding weight). We have previously shown that this food scheduling does not compromise the healthy development of young animals: the rats stay within vendor-provided weight ranges for normal growth (Stolyarova and Izquierdo, 2015). The rats were provided water ad libitum except during the hours of testing. The vivarium maintained a 12-h light/12-h dark cycle, with the temperature held constant at 22 °C. All drug treatment and behavioral testing took place between 07:00 and 09:00 h. All procedures were conducted in accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Chancellor's Animal Research Committee of UCLA.

Handling and Drug Treatment

Each rat was handled for a minimum of 10 minutes once per day for 5 consecutive days starting on PND 32, prior to drug treatment. Following handling, rats began treatment at PND 37. Injections were administered once per day for 15 consecutive days. The rats were randomly assigned to one of six groups: MPH high dose (methylphenidate hydrochloride, Sigma, St. Louis, MO; 3mg/kg, n=6; 3 male, 3 female), MPH low dose (1 mg/kg, n=4; 2 male, 2 female), FLX high dose (fluoxetine hydrochloride, Sigma, St Louis, MO; 10 mg/kg, n=6; 3 male, 3 female), FLX low dose (5 mg/kg, n=4; 2 male, 2 female), mAMPH escalating dose (d-methamphetamine

hydrochloride, Sigma, St. Louis, MO; 0.1–3.0 mg/kg s.c., increasing in 0.3mg/kg increments between days; n=6; 3 male, 3 female) and saline (n=6; 3 male, 3 female). The mAMPH group was treated until day 10 of injections and received saline for the remaining 5 days of injections. This was done to ensure the treatment reached its maximum escalating dose of 3 mg/kg to remain more consistent with the duration of treatment in our previous-published study (Ye et al., 2014). However, mAMPH treatment in the present experiment was initiated 5 d earlier than in Ye et al. (2014). MPH doses were selected to remain consistent with the range of doses previously published, which are known to produce clinically relevant levels of drug in the plasma (Crawford et al., 2011; Gerasimov et al., 2000). The order of injections was counterbalanced by rat identification number, treatment, and sex, and left/right placements of injections were rotated daily.

Behavioral Testing Apparatus

Behavioral testing was conducted in eight operant conditioning chambers (Model 80604 Lafayette Instrument Co., Lafayette, IN) that were housed within sound- and light- attenuating cubicles. Each chamber was equipped with a house light, tone generator, video camera, and LCD touchscreen opposing the pellet dispenser. The pellet dispenser delivered single 45-mg dustless precision sucrose pellets. Modified software (ABET II TOUCH) controlled touchscreen stimuli presentation, tone generation, tray- and house-light illumination and pellet dispensation.

Behavioral Testing

Pre-training

The pre-training protocol, similar to previously-published methods (Izquierdo et al., 2010; Kosheleff et al., 2012; Ochoa et al., 2015), consisted of a series of phases: Habituation, Initial Touch Training (ITT), Must Touch Training (MTT), Must Initiate Training (MIT), and Punish Incorrect Training (PT) designed to train rats to nose-poke, initiate a trial, and discriminate between stimuli.

Discrimination and Reversal Learning

Detailed methodological descriptions also appear in recent publications (Stolyarova et al., 2014; Ochoa et al., 2015; Stolyarova and Izquierdo, 2015). Rats were presented with two novel, white, equiluminescent stimuli that differed only in shape with predetermined reinforcement contingencies. The software enabled either a reward event in the form of sucrose pellet dispensation, paired with house-light illumination and an auditory feedback, as a result of nosepoking the correct stimulus, or a punishment as a result of nosepoking the incorrect stimulus; the latter was followed by a 10 s "time out" wherein rats were unable to initiate the next trial. If the rat committed an error and received a punishment, a correction trial was administered: this consisted of the same spatial (left/right) presentation of the stimulus until the rat nosepoked correctly. Spatial configuration of stimuli presentation occurred pseudorandomly, the stimulus could not have appeared on the same side of the screen more than three times in a row except during a correction trial. Stimulus assignment was counterbalanced across treatment groups. Criterion for advancement was 60 rewards at 85% correct responses within 45 min across two consecutive days. Upon reaching criterion on this phase, the rats were tested on a reversal of the

reward contingencies. In reversal, the reward contingencies were switched such that the previously unrewarded stimulus now led to a reward and the previously rewarded stimulus now led to punishment.

Tissue dissection

Rats were euthanized 9–12 days after the last day of learning (late adulthood, PND 140) with a mean of 95 days post-treatment. Rats were given an overdose of Euthasol and decapitated. The brains were immediately extracted and two millimeter-thick coronal sections of the frontal cortex, striatum, and amygdala were further rapidly dissected, using a brain matrix, over wet ice at 4°C. Frontocortical dissections included ventral (orbital) and medial sectors of the frontal cortex, but excluded most lateral, posterior (agranular insular) regions. Striatal dissections included both dorsal and ventral subregions. Following dissection, samples were immersed in isopentane (surrounded by dry ice) and then stored at −80 °C before homogenization.

ELISA method

To prepare the tissue for the assays 0.3 mL (frontal cortex, striatum) or 0.2 mL (amygdala) of PBS (0.01 mol/L, pH 7.2) containing a protease and phosphatase inhibitor cocktail (aprotinin, bestatin, E-64; leupeptin, NaF, sodium orthovanadate, sodium pyrophosphate, β-glycerophosphate; Thermo Scientific, Rockford, IL) was added to each sample. Each tissue sample was minced, homogenized, sonicated with an ultrasonic cell disrupter, and centrifuged at 5,000 g at 4°C for 10 min. Supernatants were removed and stored at −20°C until ELISA assays were performed. Bradford protein assays were also performed to determine total protein concentrations in each sample. D1R, D2R (Cat# SEB299Ra and SEA673Ra, Cloud-Clone Corp., Houston, TX)

and PSA-NCAM (Cat# 67-ABC0027B, ALPCO Diagnostics, Salem, NH) protein levels were determined using commercially-available ELISA kits. The assays were performed according to the manufacturer's instructions. The sensitivity of the assays is 0.055 ng/mL for D1R, 0.112 ng/mL for D2R, and 0.25 ng/mL for PSA-NCAM, and the detection range is 0.156–10 ng/mL for D1R, 0.312–20 ng/mL for D2R, and 0.25–16 ng/mL for PSA-NCAM. The concentration of each protein is presented as ng/mg of total protein accounting for dilution factor. Only a subset of animals (n=26) was included into D1/D2R ELISA analyses due to resource limitations. PSA-NCAM ELISAs included samples from all animals.

Data Analysis

The data were analyzed with mixed-effects generalized linear models (GLM) in MATLAB (fitglme function; Statistics and Machine Learning Toolbox; MathWorks, Natick, Massachusetts; Versions R2017a and R2020a). Sex (male vs female, categorical) and treatment group (saline, FLX, MPH vs mAMPH) were modeled as fixed factors. Analyses of sessions to chance and to criterion as well as the protein quantification data included one observation per animal. In analyzing the learning data, we focused on three main measures: probability correct, number of correction trials, and number of indiscriminate responses to the screen during ITI. For these analyses, session was included as another fixed-effects predictor and animal ID was included as a random factor. We judged the significance of each predictor based on the t-test of the associated beta coefficient. Statistical significance was noted when p-values were less than 0.05.

Across our analyses, we excluded several outlying datapoints based on visual inspection of the data. This affected PSA-NCAM and learning analyses. For PSA-NCAM, 3 samples that generated 6 measurements in total produced values that were ~3-6 times larger than the remaining

measurements. The most likely cause of these values is inadequate sample dilution or freezing conditions. The analyses within the chapter are based on the reduced sample (n=29); however, a figure showing all data, including the outliers, is included as a supplement. For the learning data, we did not exclude the data from any animal in entirely, instead 8 sessions were excluded in total across all animals and learning phases (2 sessions were excluded from 3 animals each in discrimination learning and 2 sessions were excluded from one animal in reversal learning). These sessions appear in figures but are excluded from GLM analyses. The reader may notice an example of the sessions in question in Figure 3-1A generating two unexpected drops in performance in the saline and mAMPH groups. In this case, only one animal was still learning in the saline group (with her counterparts having already passed the criterion) and her performance dropped abruptly from 70% to barely over 20%. While still presenting all of the data, we removed such data points from the GLM analyses to avoid overweighting the performance of increasingly fewer animals at the extremes.

Results

Discrimination Learning

To test the effects of adolescent drug exposure on subsequent discrimination learning in adulthood, we analyzed the behavioral data with a GLM with sex (male vs female, categorical), treatment group (saline, FLX, MPH vs mAMPH, categorical), and session (continuous) as fixed factors and rat ID as the random factor. These analyses revealed a sex difference in control animals exposed only to saline injections during adolescence, with female animals acquiring the discrimination task at a slower rate across sessions [**Figure 3-1**; β=-4.9798, t(426)=-2.7147, p=0.0069039].
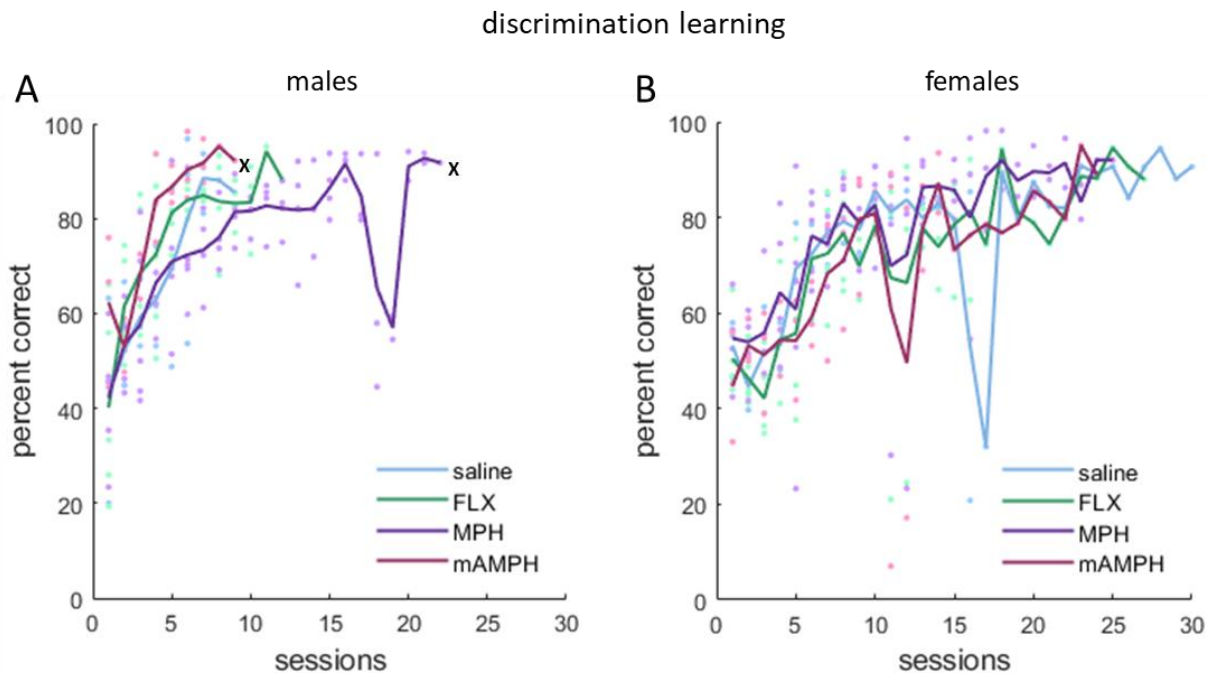
Figure 3-1. The effects of sex and adolescent drug treatment on percent correct in discrimination learning. Control females, exposed only to saline injections during the adolescent period, learned the discrimination task at a slower rate than their male counterparts. **A.** In males, mAMPH pretreatment improved the overall discrimination performance and MPH pretreatment decreased the rate of discrimination learning across sessions in adulthood. **B.** Neither of these effects were observed in females. Solid lines represent group averages and dots represent individual animal data across sessions. x<0.05.

The analyses also revealed that the effects of drug treatment depend on sex. In males, pre-exposure to mAMPH during the adolescent period improved the overall performance accuracy [β=18.95, t(426)=2.7497,p=0.0062182] and pre-exposure to MPH decreased the rate of discrimination learning across sessions [β=-5.8025, t(426)=-3.292, p= 0.0010775] in adulthood. Neither of these effects were present in females [difference between females and males in the effect of mAMPH: β=-27.115, t(426)=-3.3226, p=0.00096888; difference between females and males in the effects of MPH: β=5.3966, t(426)=2.7598, p=0.0060339).
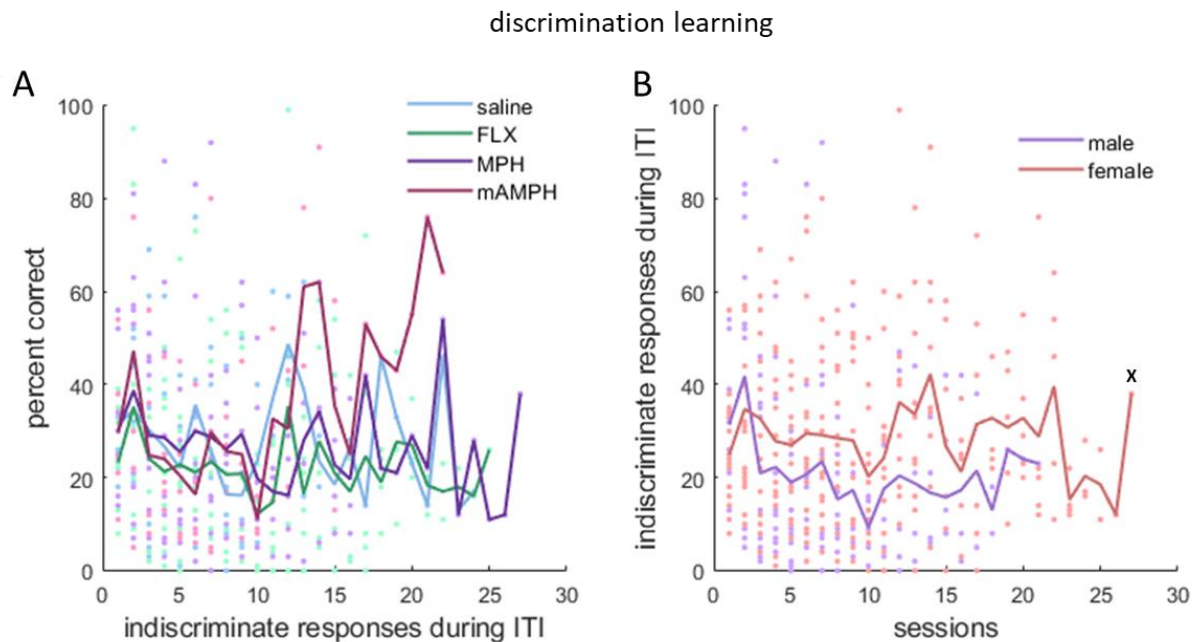
Figure 3-2. The effects of sex and adolescent drug treatment on indiscriminate responding in discrimination learning. Indiscriminate touches were defined as screen nosepokes during the ITI period, when such pokes had no programmed consequences. **A.** Indiscriminate responding decreased as animals learned the task and there were no treatment group differences. **B.** The rate of the decrease in indiscriminate responding across sessions was lower in female, compared to male, animals. Solid lines represent group averages and dots represent individual animal data across sessions. x=0.05(05).

We predicted that as animals mastered the discrimination learning task, their behavior would become more directed toward task-relevant stimuli and events and less focused on task-

irrelevant cues. We analyzed the number of indiscriminate touches that the animals made to the screen during the ITI period, when these touches had no programmed consequences. We found a marginally significant gender difference: in males, the number of indiscriminate touches decreased across sessions as expected [β=-3.636, p(394)=-2.5039, p=0.012687]; in females, however, this decrease occurred at a lower rate [difference in slope between females and males: β=2.9527, t(394)=1.9617, p=0.050505; **Figure 3-2**].
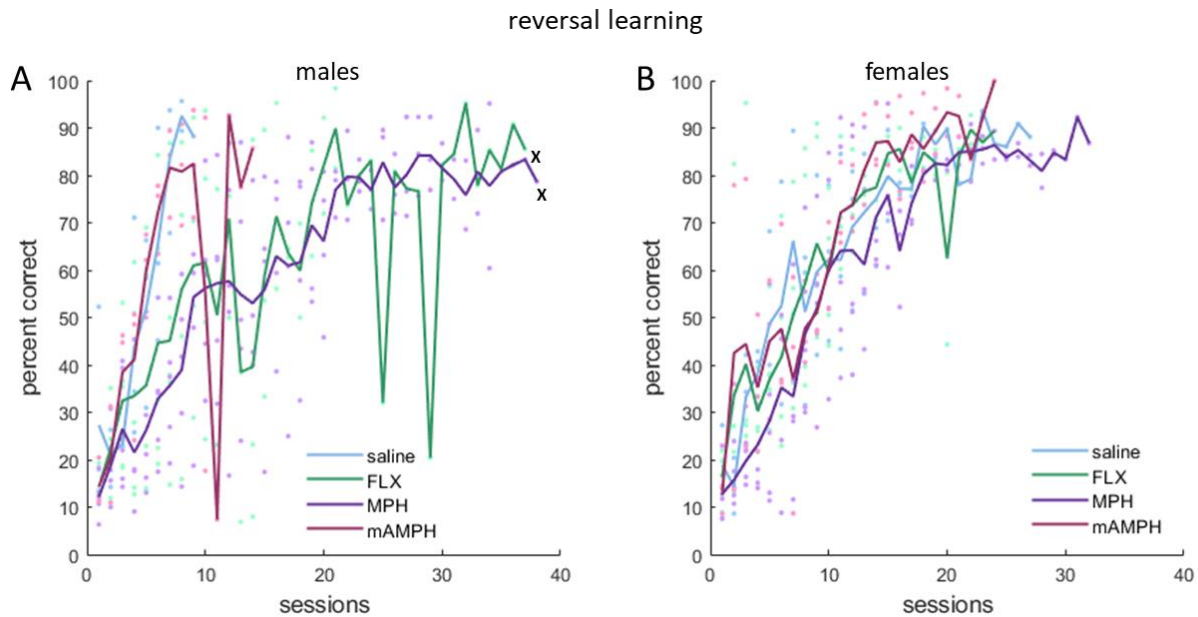
Figure 3-3. The effects of sex and adolescent drug treatment on percent correct in reversal learning. Control females, exposed only to saline injections during the adolescent period, learned the reversal task at a slower rate than their male counterparts. **A.** In males, adolescent exposure to both FLX and MPH decreased the rate of learning the reversal across sessions in adulthood. **B.** The females were protected against the deteriorating effects of these drugs; there were no treatment group differences observed for females. Solid lines represent group averages and dots represent individual animal data across sessions. x<0.05.

Reversal learning

We found that sex differences in learning persisted into reversal: while all animals improved their performance across sessions, females did so at a slower rate [**Figure 3-3**; for control males: $\beta=10.471$, $t(582)=9.5021$, $p=5.2971*10^{-20}$; difference in slope between control females and males $\beta=-6.6533$, $t(582)=-4.4148$, $p=1.2052*10^{-05}$).
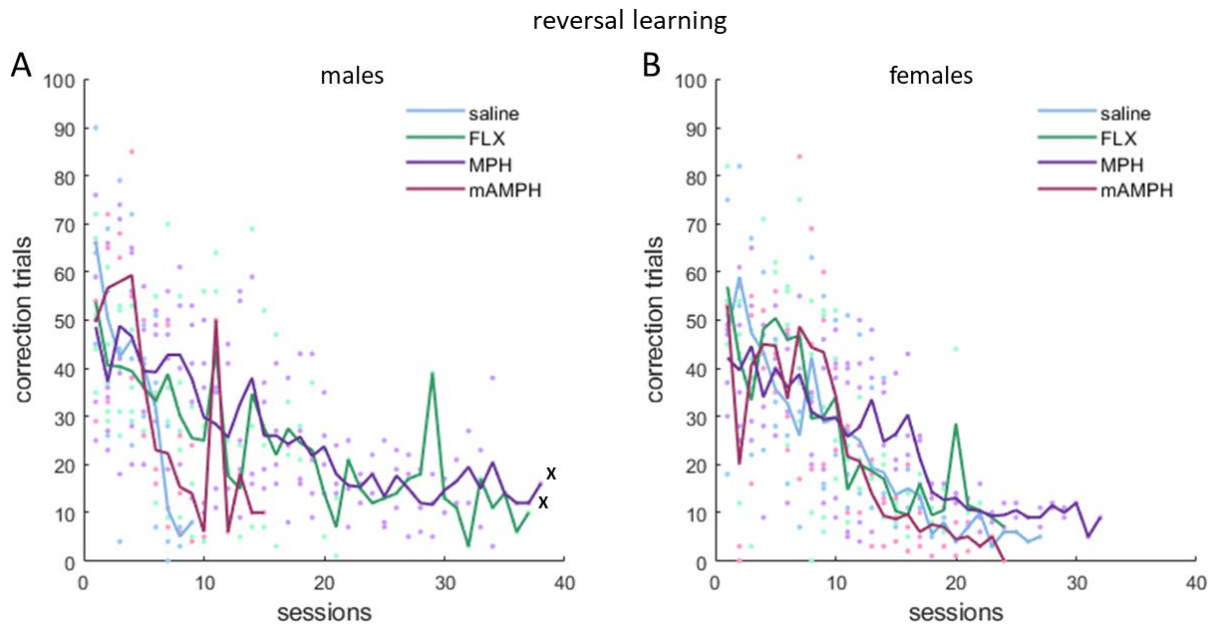
Figure 3-4. The effects of sex and adolescent drug treatment on correction trials in reversal learning. The number of correction trials can serve as an index of perseveration and outcome-insensitive responding. **A.** In males, adolescent exposure to both FLX and MPH reduced the rate of the decrease in correction trials across sessions of the reversal. **B.** The females were protected against the deteriorating effects of these drugs; there were no treatment group differences observed for females. Solid lines represent group averages and dots represent individual animal data across sessions. x<0.05.

The effects of drug treatment on performance accuracy (percent correct) during reversal learning also varied with sex. We found that in males pretreatment with both FLX [$\beta=-4.5302$, $t(582)=-2.4693$, $p=0.013822$] and MPH [$\beta=-7.6441$, $t(582)=-5.2644$, $p=1.9819*10^{-07}$] during the adolescent period decreased the rate of reversal learning across sessions in adulthood (**Figure 3-**

3). Neither of these drugs had a deteriorating effect on reversal learning performance in females [difference in slopes between females and males, FLX: β=5.9558, t(582)=2.5059, p=0.012486; MPH: β=8.498, t(582)=4.2228, p=2.799*10^{-05}].
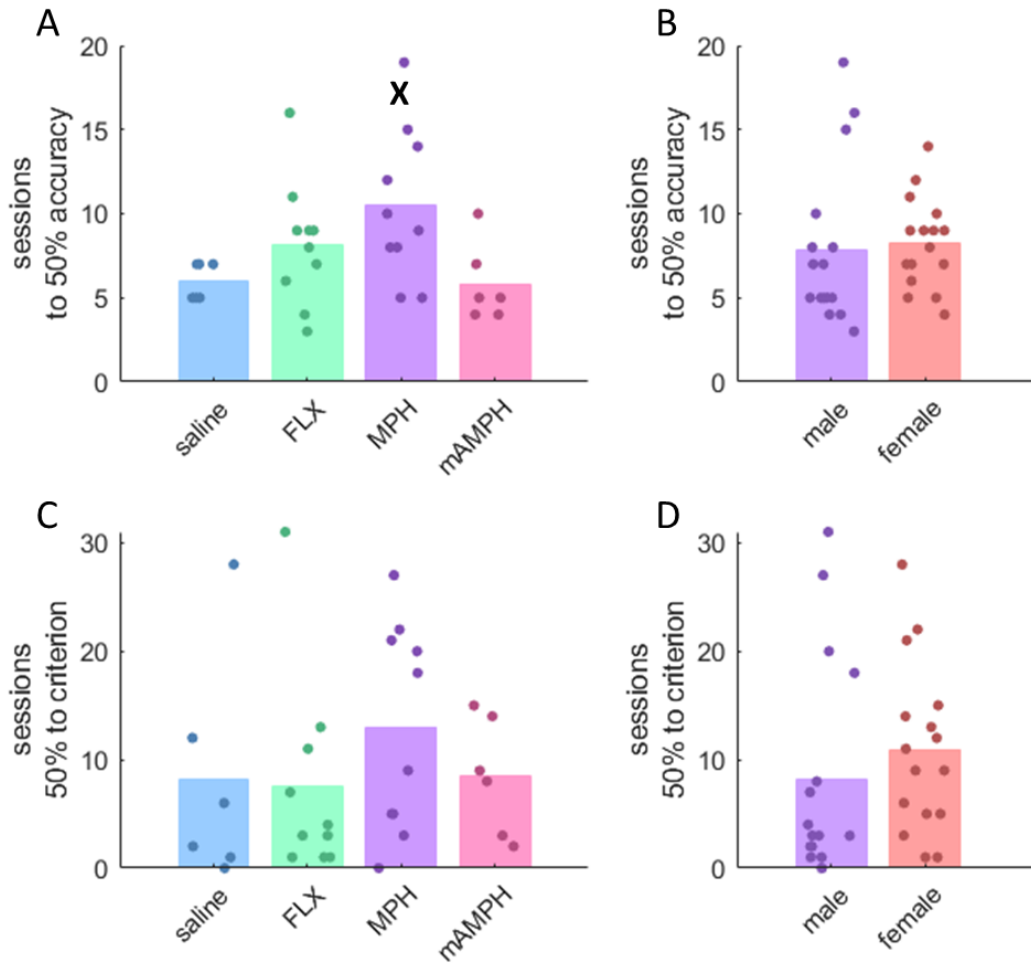
Figure 3-5. The effects of sex and adolescent drug treatment on sessions to chance and criterion in reversal learning. A. Only MPH pretreatment increased the number of sessions to the 'at-chance' level of performance. **B.** There were no sex differences in the number of sessions to the 'at-chance' level of performance. **C-D.** Neither drug pretreatment, nor sex affected the number of sessions to reach the criterion after the animals have overcome the chance level of performance. Bars represent group averages and dots represent individual animal data. x<0.05.

We also analyzed the number of correction trials, which serve as an index of perseveration and outcome-insensitive response repetition. While the number of correction trials decreased across sessions, the rate of the decrease was lower in males exposed to FLX [β=4.4058, $t(581)$=2.196, p=0.028487] and MPH [β=6.6792, $t(581)$=3.9093, p=0.00010349] during the adolescent period compared to their saline counterparts (**Figure 3-4**). Females, on the other hand, were protected against these negative effects of drug exposure [difference in slopes between females and males, FLX: β= -4.7421, $t(581)$=-2.181, p=0.029583; MPH: β=-6.1152, $t(581)$=-3.28, p=0.0011003].

Finally, we tested whether pretreatment with FLX and MPH specifically affected early reversal learning before animals reached the chance level of performance, as this stage is characterized by the greatest demands to overcome perseveration. Only pretreatment with MPH increased the total number of sessions to reach the 'at-chance' level of performance across male and female animals combined [**Figure 3-5**; β=2.8667, $t(24)$=3.1581, p=0.00425]. None of the drugs affected the number of sessions that the animals required to reach the reversal performance criterion after overcoming the 'at-chance' level (all p values > 0.08).

PSA-NCAM, D1 and D1 receptor expression

We quantified the expression of PSA-NCAM in the frontal cortex and amygdala and D1 and D2 receptors in the striatum and analyzed these measures with a GLM with sex (male vs female, categorical) and treatment group (saline, FLX, MPH vs mAMPH, categorical) as fixed factors. We found that PSA-NCAM expression in the cortex was increased in male, compared to female, rats [**Figure 3-6**; β=1.3521, $t(21)$=4.8319, p=8.9217*10^{-05}]. It was, however, unaffected

by adolescent drug exposure (all p values > 0.34). In the amygdala, PSA-NCAM expression was affected by neither the sex of the animal nor adolescent drug treatment (all p values > 0.2).
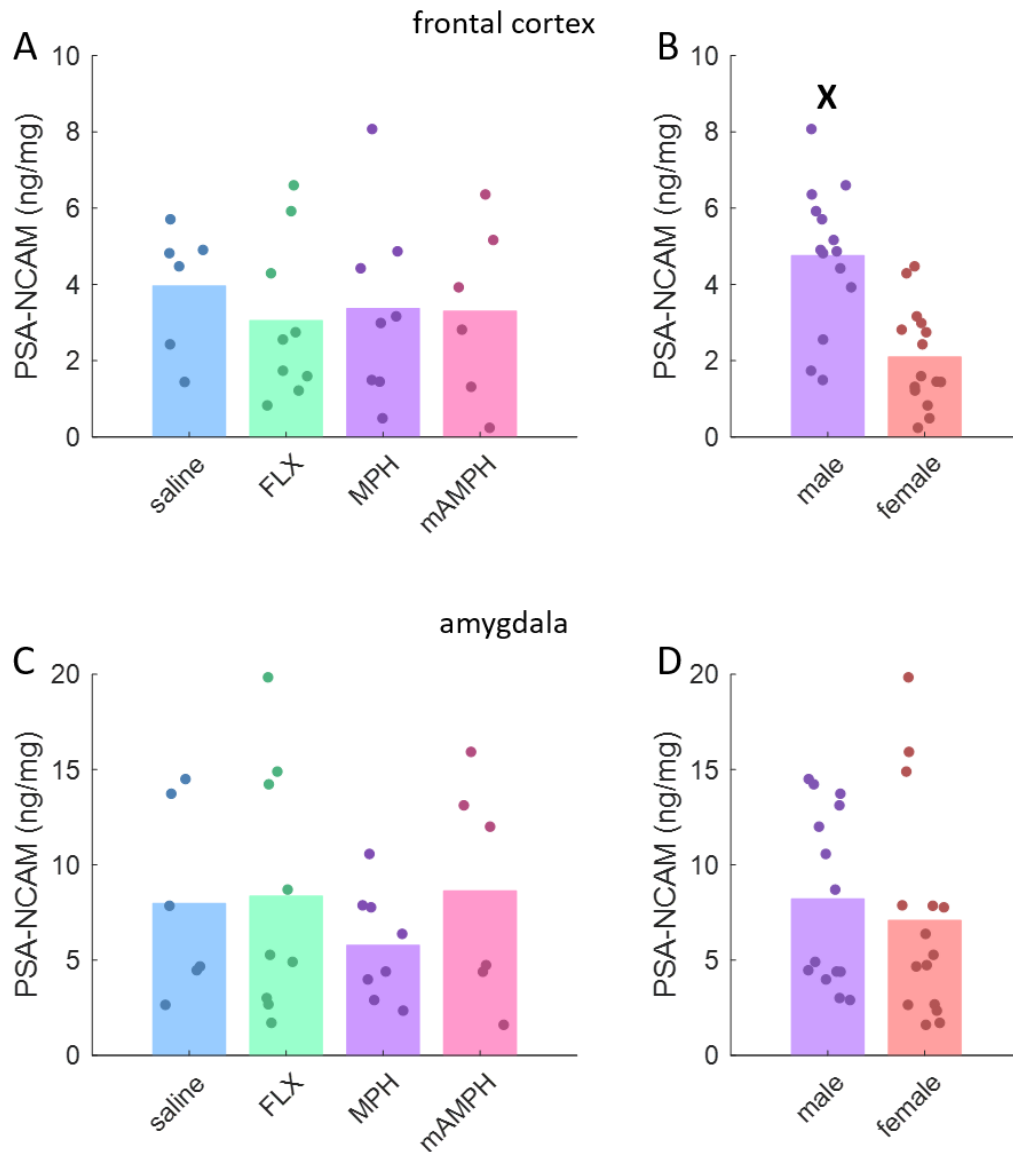
Figure 3-6. The effects of sex and adolescent drug treatment on PSA-NCAM expression in the frontal cortex and amygdala. A. Cortical expression of PSA-NCAM was unaffected by the adolescent drug treatment. **B.** The levels of cortical PSA-NCAM expression were higher in male than in female animals. **C-D.** Neither the drug pretreatment, nor sex affected the PSA-NCAM expression in the amygdala. Bars represent group averages and dots represent individual animal data. x<0.05. See *Figure 3-Supp1* at the end of this chapter for the representation of outliers.
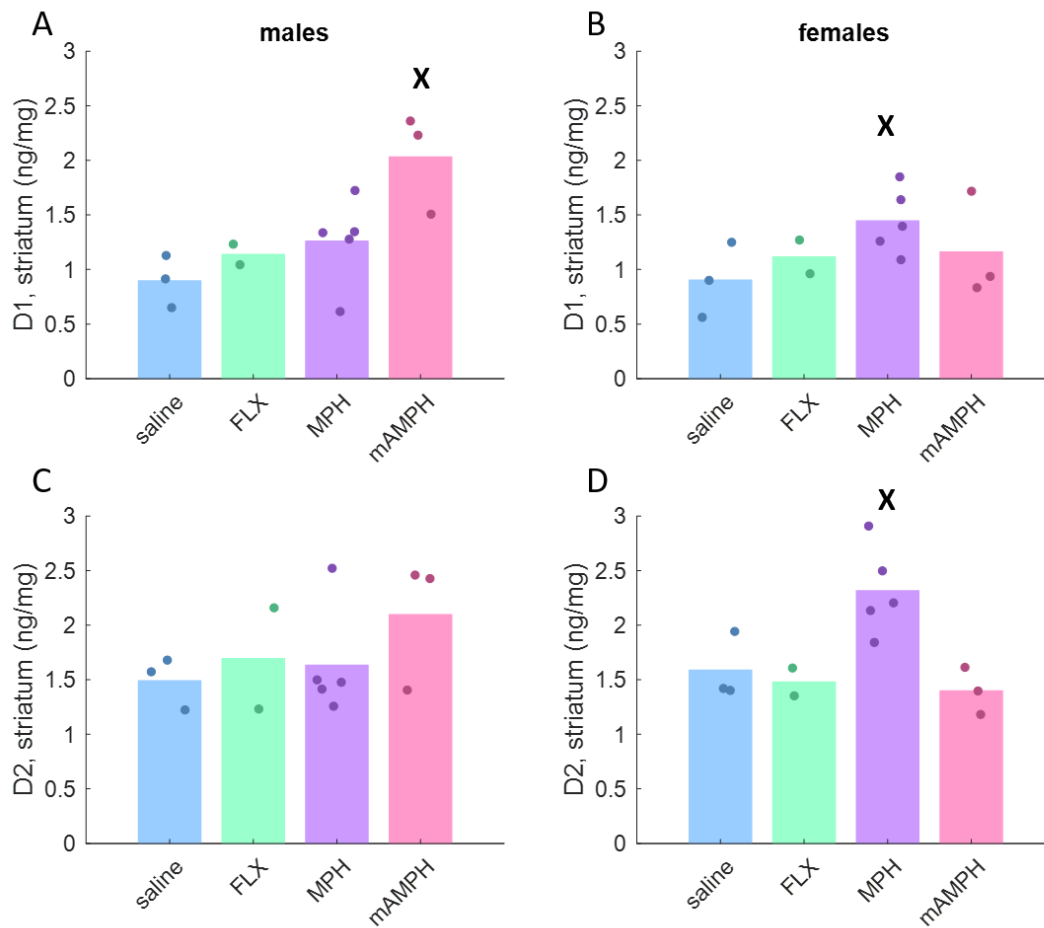
Figure 3-7. The effects of sex and adolescent drug treatment on D1 and D2 receptor expression in the striatum. A, B. D1 receptors were upregulated in the striatum of adult male animals after adolescent mAMPH exposure and of adult female animals after adolescent MPH exposure. **C, D.** Striatal D2 receptors were significantly upregulated in females, but not males, after the adolescent MPH treatment. Bars represent group averages and dots represent individual animal data. x<0.05.

Our analysis of striatal D1 receptor expression revealed a significant sex x treatment interaction for the effect of mAMPH [**Figure 3-7**; β=0.3478, t(18)=3.2818, p=0.0041447] and a marginally significant sex x treatment interaction for the effects of MPH [β=-0.18088, t(18)=-1.99, p=0.061997]. Post hoc analyses revealed that the expression of D1 receptors in the striatum was increased after mAMPH treatment in males [β=0.70088, t(9)=4.5926, p=0.0013042], but not in females [β=0.0052731, t(9)=0.035848, p=0.97219]. Striatal D1 receptors were also upregulated

after adolescent MPH treatment in female [**Figure 3-7**; β=0.2896, t(9)=2.2956, p=0.047341], but not male [β=-0.072161, t(9)=-0.55133, p=0.59483], rats. For striatal D2 receptors, we founds significant sex x treatment interactions for the effects of mAMPH [β=0.33298, t(18)=2.6521, p=0.016215] and MPH [β=-0.35876, t(18)=-3.3318, p=0.0037113]. Post hoc analyses revealed no consistent effects of mAMPH treatment for either males [β=0.36774, t(9)=1.7476, p=0.11447] or females [β=-0.29822, t(9)=-2.1765, p=0.057505]; as such, these results are difficult to interpret. For MPH, we found that D2 receptors in the striatum were significantly upregulated in females [β=0.62163, t(9)=5.2899, p=0.0005005], but not males [β=-0.095888, t(9)=-0.53133, p=0.60805], after adolescent MPH pretreatment.
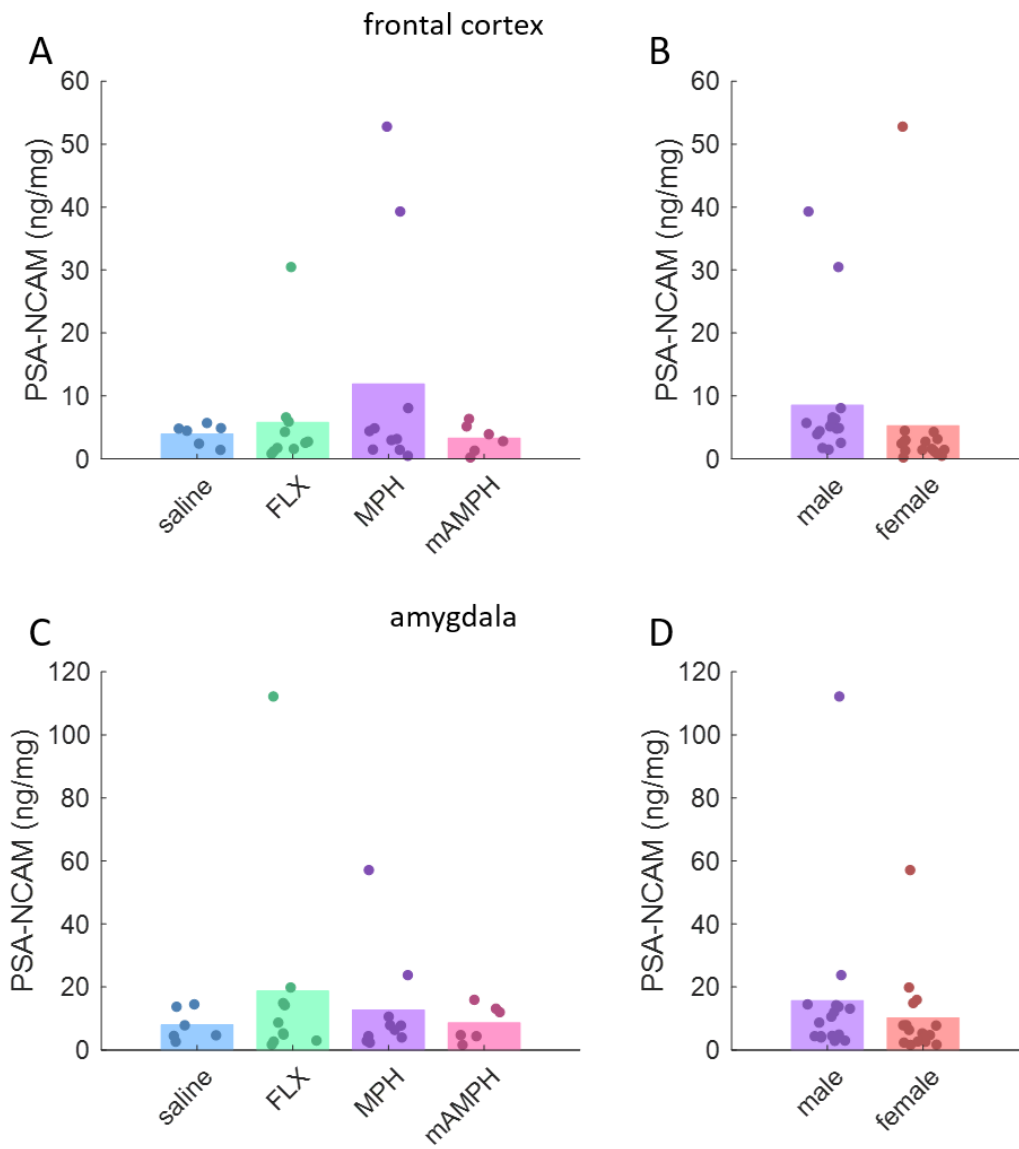
Figure 3-Supp1. The outliers in analyses of PSA-NCAM expression in the frontal cortex and amygdala. Across our analyses we excluded 6 outlying datapoints based on visual inspection of the data: 3 samples that generated 6 measurements in total produced values that were ~3-6 larger than the remaining measurements. The most likely cause of these values is inadequate sample dilution or freezing conditions. Bars represent group averages and dots represent individual animal data.

Discussion

Learning and cognitive flexibility in adulthood after adolescent drug exposure

We report the novel finding that adolescent exposure to MPH and FLX has long-lasting consequences for flexible reward learning in adulthood. In our study, adolescent exposure to MPH produced the most robust impairment in male rats. These males were slower to acquire discrimination and reversal learning and demonstrated more perseverative responding compared to their saline counterparts. Acutely, MPH has been shown to enhance cognition by improving attention and memory performance (Mehta et al., 2004). In our study rats underwent a washout period followed by extensive off-drug testing on the discrimination learning task before undergoing reversal learning, – the timeframe during which the impairment was observed.

Previous studies have reported that MPH increases impulsivity in rats prescreened to be 'low-impulsive' before exposure (Caprioli et al., 2015). Accordingly, a subset of our adult rats pretreated with MPH in adolescence may have had difficulty attending to the task demands in reversal learning. In support of this, we demonstrate slower early reversal learning (i.e., "below chance") when the levels of choice ambiguity and cognitive demand are high. This is similar to the long-lasting effects of mAMPH in adulthood (Stolyarova et al., 2014). In that study, we found an impairment in overcoming initial perseveration, but unaltered performance once the animals overcame the "at chance" level.

Adolescent FLX exposure produced a similar pattern of impairment, albeit only in reversal learning. Male rats exposed to FLX during the adolescent period acquired reversal learning at a slower rate compared to their saline counterparts. Previous studies have shown developmental FLX to engender an anxiogenic profile in adulthood (Iñiguez et al., 2010; Iñiguez et al., 2014; Homberg et al., 2011) and a cocaine preference later in life (Iñiguez et al., 2015). Reduced learning

flexibility in this group could be due to increased stress reactivity and a compromised ability to cope with changes in new task demands. Recent evidence shows that the long-term effects may be different among different SSRIs: paroxetine, citalopram, and FLX (Schaefer et al., 2013; Altieri et al., 2014; Amodeo et al., 2015); and furthermore, that the long-term effects depend on the age of exposure and whether the animals are later tested on or off drug (Homberg et al., 2011; Vorhees et al., 2011). FLX exposure in adulthood rather than adolescence instead facilitates early reversal learning in mice tested on a touchscreen paradigm similar to the present task wherein animals are presented with a concurrent pairwise discrimination problem (Brigman et al., 2010). To add to the complexity, the engagement of the serotonin system may depend on the parameters of the task: for example, probabilistic tasks involving uncertainty about stimulus-reward contingencies may be more reliant on finely tuned serotonergic modulation, compared to deterministic tasks (Rygula et al., 2014; Ochoa et al., 2015).

Surprisingly, mAMPH pretreated rats did not display the discrimination or reversal learning impairments we previously observed when animals were treated in later adolescence, beginning in PND 41 (Ye et al., 2014). Instead, we observed an improvement in discrimination learning after adolescent mAMPH treatment in male rats. There is now evidence that there are differences in reward sensitivity and addiction vulnerability in early vs. late rat adolescence and that these effects are often sex dependent (Spear, 2015). There may also be a difference in the aversive properties of mAMPH depending on age of exposure: in one study, a 9 d difference in adolescence had a significant impact on drug response (Vorhees et al., 2011).

The enhanced discrimination learning on a reward-based task in protracted withdrawal from mAMPH may seem to be at odds with previous literature on reward-deficiency in addiction (Hommer et al., 2011; Destoop et al., 2019; Blum et al., 2012). We suggest that an animal's

response to palatable rewards and its ability to learn about them depend on whether drug-taking is one of the behavioral options in the environment. It has been previously proposed that the value of an option is sensitive to the overall reinforcement rate or reward availability in the environment. For example, when an individual has frequent experience with a drug of abuse, the average rate of reinforcement is inflated and the benefits conferred by natural rewards pale in comparison (Dezfouli et al., 2009). We argue that the situation is reversed during protracted drug abstinence, when individuals are confronted with a relatively reward-poor environment in comparison to the previous on-drug state. When the average rate of rewards is low, the impact of each reinforcing experience is magnified, driving more rapid learning

Sex differences

Among the control animals (those that only received saline during adolescence), female rats learned both the initial discrimination and subsequent reversal at a slower rate than their male counterparts. To our knowledge, a sex difference in visual discrimination and reversal learning in untreated rats has not been reported before. Sex differences in previous studies have been noted in the areas of increased addiction vulnerability (Crawford et al., 2011), higher levels of anxiety (Iñiguez et al., 2010), and higher levels of impulsivity in male vs. female rats (Caprioli et al., 2015). Rodent females, like human females, mature earlier than males: on average the development of genitalia and activation of sexual organs occurs 4–8 d sooner (Spear, 2015). In addition to this difference, hormone signaling and pharmacokinetic differences between the sexes (Shanks et al., 2015; Crawford et al., 2011) may also have contributed to the sex differences we report here. It is possible that the estrous cycle produced some effects on learning. Current conventional methods to assess the estrous cycle in intact females require obtaining vaginal

smears, which would have introduced stress as a variable. Since there is evidence that even brief stressors affect learning flexibility (Izquierdo et al., 2006; George et al., 2015), we did not perform this assessment in the present experiment. Future, adequately-powered experiments should systematically examine the relationship of hormonal influences on these measures.

Notably, we only found effects of drug treatment in male rats. The females in the present study were not affected by drug treatment during the adolescent period: they did not show the same long-lasting changes in learning and cognitive flexibility as males did. It should be noted that drug exposure using animal models of depression (Iñiguez et al., 2014) or ADHD (Vendruscolo et al., 2008; Baskin et al., 2015) may have yielded different results. Most groups, however, treat typical, normally developing animals (as in the present study).

PSA-NCAM expression after adolescent psychostimulant exposure

PSA-NCAM modulates synaptic remodeling and plasticity, partly by supporting the formation of new synapses (Muller et al., 1996, 2010; Durbec and Cremer, 2001; Vutskits et al., 2001). Its expression increases during adolescence and may contribute to functional reorganization within the frontocortical, hippocampal, amygdalar and striatal circuits (Nacher et al., 2002a,b; Seki, 2002; Varea et al., 2005). We have previously found that expression of PSA-NCAM in the frontal cortex and amygdala is increased in animals that went through cognitive training and reward learning in adolescence compared to animals that had similar experiences in adulthood. Furthermore, levels of PSA-NCAM are predictive of willingness to invest effort in pursuit of desirable rewards, indicative of a reward sensitive endophenotype (**Chapter 2**). In the present study we assessed PSA-NCAM expression levels after adolescent psychostimulant exposure and prolonged discrimination and reversal training. We found no effects of drug treatment on PSA-

NCAM levels. We did, however, observe a significant effect of sex, with females having lower PSA-NCAM levels in the frontal cortex. This effect of sex has not been previously reported to our knowledge and highlights the need for research into sex differences in neurodevelopmental processes. Because PSA-NCAM expression has been previously shown to be sensitive to reward and learning experiences (Pham et al., 2003; Cordero et al., 2005; Bisaz et al., 2009; Stolyarova and Izquierdo, 2015), we think it is likely that adolescent drug exposure may induce some changes in its levels, but not at the time point at which it was assessed in the present study. Firstly, there was a large temporal gap between the termination of treatment and PSA-NCAM quantification. Secondly, PSA-NCAM levels tend to decrease significantly with maturity and aging: animals in the current study were at PND 140 at the time of brain collection, while our animals in the previous study were much younger (PND 50 for late adolescents and PND 86 for adults).

Striatal D1 and D2 receptor expression after adolescent psychostimulant exposure

Corticostriatal circuitry is critical to adaptive learning and motivation (Cagniard et al., 2006). D1 and D2 receptor signaling in the striatum regulates overlapping yet dissociable aspects of reward learning and decision-making (Izquierdo et al., 2006; Schweimer and Hauber, 2006; Groman et al., 2011; Stopper et al., 2013; Keeler et al., 2014; Yohn et al., 2015). The adolescent period is marked by the reduction in density of both types of receptors in the striatum (Gelbard et al., 1989; Teicher et al., 1995; Tarazi and Baldessarini, 2000). In the present investigation, we found significantly increased D1 receptor expression in adult male learners that were previously exposed to mAMPH during the adolescent period and in adult female learners that were previously exposure to MPH during the adolescent period. Interestingly, D1-mediated signaling has been previously linked to the behavioral phenotype of a rat model of ADHD (Ohno et al., 2012). The

ability of MPH treatment to alter D1 receptor expression following learning may contribute to its clinical efficacy in the adolescent population. Increased striatal D1 receptor expression (and by extension, binding) would result in enhanced excitability of the pathway involved in learning about and responding to rewards (Cox et al., 2015). This result is consistent with what has been found for the role of D1 in the prefrontal cortex: enhanced D1 function in the prefrontal cortex predicts general cognitive abilities (Wass et al., 2013). Similarly, developmental psychostimulant exposure may interact with later experience with reward (in discrimination and reversal learning) to upregulate D1 expression in the striatum, leading to an enduring reward-sensitive phenotype. We note here that this important 'later experience' could simply be more environmental enrichment: it may not be the (food) reward exposure or learning per se, but rather the more complex environment and increased option space that crucially engage DA signaling at that later timepoint. In order to ascertain that such receptor expression changes are due to reward learning experience and not due simply to maturational changes, an appropriate age-matched homecage control group should be added to future investigations.

We propose here, as have others previously, that reduced or excessive (supranormal) DA activity can have different effects on cognitive processes, depending on region-specific receptor activation (Floresco, 2013). For example, local infusions of MPH in the (baso)lateral amygdala (BLA) enhance cue-reward learning through a D1 mechanism and suppress task-irrelevant behaviors through a D2 mechanism (Tye et al., 2010; Larkin et al., 2015). Therefore, chronic administration of MPH may result in striatal downregulation of D1 receptors in the short term, but when assessed after prolonged drug-withdrawal and upon conditions of reward learning, D1 receptors may be upregulated in the long-term. We only assessed D1 and D2 receptor expression in the striatum in the present study, however DA signaling may be affected differently in another

region where DA expression may correlate meaningfully with learning, such as the BLA (Wassum and Izquierdo, 2015). Additionally, since we collected striatal samples that included both dorsal and ventral regions of the striatum, it is possible that subregional differences may have masked receptor expression effects.

We previously found that D1 expression decreases in adolescent animals that had prior experience with (food) reward learning, compared to animals that had the same experience in adulthood (Stolyarova and Izquierdo, 2015). Decreased D1 expression at the onset of adulthood is predicted to render animals less reward sensitive, whereas increased D1 expression in adolescence may help to establish this reward-sensitive phenotype. Of course, there are differences in measures of DA receptor expression, availability, and binding, and with our current methods, we are unable to detect differences in functionality. For example, there is the possibility that there may be D1 turnover changes, trafficking and insertion of receptors "on demand," or silent synapses (Dong and Nestler, 2014) that we are unable to capture with our protein assays. However, protein expression assessment via ELISA provides an advantage over binding studies, as it allows the distinction between D1 and D5 subpopulations of the D1-like family of receptors. Taken together with results from untreated adolescent vs. adult animals (Stolyarova and Izquierdo, 2015), our findings are consistent with the 'prepare and select' model of striatal D1 and D2 receptors, respectively (Keeler et al., 2014). Increased striatal D1 receptor expression and/or availability would be expected to engender a readiness to respond to reward in animals pretreated with MPH or mAMPH. It is noteworthy that food restriction on its own has been previously shown to upregulate D1 receptors in the ventral striatum (Carr et al., 2009). Since all of the groups in the present study experienced identical food restriction conditions, MPH and mAMPH pretreatment

during adolescence may have rendered striatal D1 transcription machinery more responsive to environmental changes.

Two months of treatment with MPH beginning in PND 30, similar to our timepoint, results in decreases in D2 availability assessed with in vivo microPET (Thanos et al., 2007), suggesting an addiction-vulnerable profile. Ontogenetic changes in D2 receptors may be partially responsible for differences in psychostimulant sensitivity (McDougall et al., 2015) since the functionality of the D2 receptor continues to mature beyond the preweanling period (Der-Ghazarian et al., 2014) and likely through adolescence. Other groups, assessing D2 receptor density via autoradiography after early MPH exposure (PND 21) do not report enduring effects of the drug on either D1 or D2 receptor density (Gill et al., 2013). However, MPH may have long lasting effects on the functionality and expression of D2 receptors depending on early vs. late adolescent exposure. In the present study, we found that striatal D2 receptors were upregulated in female, but not male, adults previously exposed to MPH during the adolescent period. Sex differences in responses to developmental psychostimulant exposure and the associated changes in the DA D1 and D2 receptor expression should be explored further. If lower D2 expression is associated with addiction vulnerability, our results suggest that the juvenile MPH exposure may have efficacy in reducing this risk in females. Lastly, an important area for investigation is to determine how the pharmacodynamics and plasma half-lives of MPH and mAMPH may differentially contribute to the selective receptor expression and learning effects.

Conclusions

In the present report, we show sex differences in visual discrimination and reversal learning, assessed in a novel environment. Though the estrous cycle was not measured in the current experiment, our data provide a basis for future systematic inquiry into sex differences in reward learning and cognitive flexibility. We also report the first evidence of enduring effects of adolescent MPH and FLX exposure on reversal learning in adulthood that is specific to male rats. These findings have limited implications for learning flexibility and adaptive decision-making in a human clinical population (those diagnosed with ADHD), but may have the most relevance to an adolescent recreational user population. To that end, our results show that developmental psychostimulant exposure may interact with reward experience to boost D1 and D2 receptor expression in a sex-dependent manner later in life. This may be particularly analogous to the young human recreational user that consumes psychostimulants as cognitive enhancers to boost academic performance.

References

Achat-Mendes, C., Anderson, K. L., Itzhak, Y. (2003). Methylphenidate and MDMA adolescent exposure in mice: long-lasting consequences on cocaine-induced reward and psychomotor stimulation in adulthood. *Neuropharmacology* 45(1), 106–115.

Adriani, W., Leo, D., Greco, D., Rea, M., di Porzio, U., Laviola, G., et al. (2006). Methylphenidate administration to adolescent rats determines plastic changes on reward-related behavior and striatal gene expression. *Neuropsychopharmacology* 31(9), 1946–1956.

Altieri, S. C., Yang, H., O'Brien, H. J., Redwine, H. M., Senturk, D., Hensler, J. G., et al. (2015). Perinatal vs. genetic programming of serotonin states associated with anxiety. *Neuropsychopharmacology* 40(6), 1456–1470.

Amodeo, L. R., Greenfield, V. Y., Humphrey, D. E., Varela, V., Pipkin, J. A., Eaton, E., et al. (2015). Effects of acute or repeated paroxetine and fluoxetine treatment on affective behavior in male and female adolescent rats. *Psychopharmacology* 232(19), 3515-3528.

Baskin, B. M., Dwoskin, L. P., Kantak, K. M. (2015). Methylphenidate treatment beyond adolescence maintains increased cocaine self-administration in the spontaneously hypertensive rat model of attention deficit/hyperactivity disorder. *Pharmacol Biochem Behav* 131, 51–56.

Bisaz, R., Conboy, L., and Sandi, C. (2009). Learning under stress: a role for the neural cell adhesion molecule NCAM. *Neurobiol Learn Mem* 91, 333–342.

Blum K., Gardner E., Oscar-Berman M., Gold M. (2012). "Liking" and "wanting" linked to Reward Deficiency Syndrome (RDS): hypothesizing differential responsivity in brain reward circuitry. *Current Pharmaceutical Design* 18, 113–118.

Brigman, J. L., Mathur, P., Harvey-White, J., Izquierdo, A., Saksida, L. M., Bussey, T. J., et al. (2010). Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. *Cerebral Cortex* 20(8), 1955–1963.

Cagniard, B., Beeler, J. A., Britt, J. P., McGehee, D. S., Marinelli, M., Zhuang, X. (2006). Dopamine scales performance in the absence of new learning. *Neuron* 51(5), 541–7.

Caprioli, D., Jupp, B., Hong, Y. T., Sawiak, S. J., Ferrari, V., Wharton, L., et al. (2015). Dissociable Rate-Dependent Effects of Oral Methylphenidate on Impulsivity and D2/3 Receptor Availability in the Striatum. *The Journal of Neuroscience* 35, 3747–3755.

Carr, K. D., Cabeza de Vaca, S., Sun, Y., Chau, L. S. (2009). Reward-potentiating effects of D-1 dopamine receptor agonist and AMPAR GluR1 antagonist in nucleus accumbens shell and their modulation by food restriction. *Psychopharmacology (Berl)* 202(4), 731–743.

Castillo-Gómez, E., Gómez-Climent, M. A., Varea, E., Guirado, R., Blasco-Ibáñez, J. M., Crespo, C., et al. (2008). Dopamine acting through D2 receptors modulates the expression of PSA-NCAM,

a molecule related to neuronal structural plasticity, in the medial prefrontal cortex of adult rats. *Exp Neurol* 214, 97–111.

Cordero, M. I., Rodríguez, J. J., Davies, H. A., Peddie, C. J., Sandi, C., and Stewart, M. G. (2005). Chronic restraint stress down-regulates amygdaloid expression of polysialylated neural cell adhesion molecule. *Neuroscience* 133, 903–910.

Cox, S. M., Frank, M. J., Larcher, K., Fellows, L. K., Clark, C. A., Leyton, M., et al. (2015). Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *Neuroimage* 109, 95–101.

Crawford, C. A., Baella, S. A., Farley, C. M., Herbert, M. S., Horn, L. R., Campbell, R. H., et al. (2011). Early methylphenidate exposure enhances cocaine self-administration but not cocaine-induced conditioned place preference in young adult rats. *Psychopharmacology* 213, 43–52.

Destoop, M., Morrens, M., Coppens, V., Dom, G. (2019). Addiction, Anhedonia, and Comorbid Mood Disorder. A Narrative Review. *Front Psychiatry* 10, 311.

Der-Ghazarian, T., Widarma, C. B., Gutierrez, A., Amodeo, L. R., Valentine, J. M., Humphrey, D. E., et al. (2014). Behavioral effects of dopamine receptor inactivation in the caudate-putamen of preweanling rats: role of the D2 receptor. *Psychopharmacology* 231(4), 651–662.

Dezfouli, A., Piray, P., Keramati, M. M., Ekhtiari, H., Lucas, C., Mokri, A. (2009). A neurocomputational model for cocaine addiction. *Neural Comput* 21(10), 2869-2893.

Dong, Y., Nestler, E. J. (2014). The neural rejuvenation hypothesis of cocaine addiction. *Trends Pharmacol Sci* 35(8), 374–383.

Durbec, P., and Cremer, H. (2001). Revisiting the function of PSA-NCAM in the nervous system. *Mol Neurobiol* 24, 53–64.

Floresco, S. B. (2013). Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. *Front Neurosci* 7, 62.

Gelbard, H. A., Teicher, M. H., Faedda, G., Baldessarini, R. J. (1989). Postnatal development of dopamine D1 and D2 receptor sites in rat striatum. *Brain Res Dev Brain Res.* 49, 123–130.

George, S. A., Rodriguez-Santiago, M., Riley, J., Abelson, J. L., Floresco, S. B., Liberzon, I. (2015). Alterations in cognitive flexibility in a rat model of post-traumatic stress disorder. *Behav Brain Res* 286, 256–64.

Gerasimov, M. R., Franceschi, M., Volkow, N. D., Gifford, A., Gatley, S. J., Marsteller, D., et al. (2000). Comparison between intraperitoneal and oral methylphenidate administration: A microdialysis and locomotor activity study. *J Pharmacol Exp Ther* 295(1), 51–57.

Gill, K. E., Beveridge, T. J. R., Smith, H. R., Porrino, L. J. (2013). The effects of rearing environment and chronic methylphenidate administration on behavior and dopamine receptors in adolescent rats. *Brain Research* 1527, 67–78.

Gill, K. E., Chappell, A. M., Beveridge, T. J. R., Porrino, L. J., Weiner, J. L. (2014). Chronic Methylphenidate Treatment During Early Life Is Associated with Greater Ethanol Intake in Socially Isolated Rats. *Alcoholism: Clinical and Experimental Research* 38, 2260–2268

Gray, J. D., Punsoni, M., Tabori, N. E., Melton, J. T., Fanslow, V., Ward, M. J., et al. (2007). Methylphenidate Administration to Juvenile Rats Alters Brain Areas Involved in Cognition, Motivated Behaviors, Appetite, and Stress. *J Neurosci* 27, 7196–7207.

Groman, S. M., Lee, B., London, E. D., Mandelkern, M. A., James, A. S., Feiler, K., et al. (2011). Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. *J Neurosci* 31(20), 7291–9.

Grund, T., Teuchert-Noodt, G., Busche, A., Neddens, J., Brummelte, S., Moll, G. H., et al. (2007). Administration of oral methylphenidate during adolescence prevents suppressive development of dopamine projections into prefrontal cortex and amygdala after an early pharmacological challenge in gerbils. *Brain Res* 1176, 124–132.

Homberg, J. R., Olivier, J. D., Blom, T., Arentsen, T., van Brunschot, C., Schipper, P., et al. (2011). Fluoxetine exerts age-dependent effects on behavior and amygdala neuroplasticity in the rat. *PLoS One* 6(1), e16646.

Hommer D. W., Bjork J. M., Gilman J. M. (2011). Imaging brain response to reward in addictive disorders. *Annals of the New York Academy of Sciences.* 1216, 50–61.

Iñiguez, S. D., Alcantara, L. F., Warren, B. L., Riggs, L. M., Parise, E. M., Vialou, V., et al. (2014). Fluoxetine Exposure during Adolescence Alters Responses to Aversive Stimuli in Adulthood. *J Neurosci* 34, 1007–1021.

Iñiguez, S. D., Riggs, L. M., Nieto, S. J., Dayrit, G., Zamora, N. N., Shawhan, K. L., et al. (2014). Social defeat stress induces a depression-like phenotype in adolescent male c57BL/6 mice. *Stress* 17(3), 247–255.

Iñiguez, S. D., Riggs, L. M., Nieto, S. J., Wright, K. N., Zamora, N. N., Cruz, B., et al. (2015). Fluoxetine exposure during adolescence increases preference for cocaine in adulthood. *Sci Rep* 5:15009.

Iñiguez, S. D., Warren, B. L., Bolaños-Guzman, C. A. (2010). Short- and Long-Term Functional Consequences of Fluoxetine Exposure During Adolescence in Male Rats. *Biological Psychiatry* 67, 1057–1066.

Izquierdo, A., Belcher, A. M., Scott, L., Cazares, V. A., Chen, J., O'Dell, S. J., et al. (2010). Reversal-specific learning impairments after a binge regimen of methamphetamine in rats: possible involvement of striatal dopamine. *Neuropsychopharmacology* 25(2), 505–514.

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., Holmes, A. (2016). The neural basis of reversal learning: An updated perspective. *Neuroscience* 345, 12-26

Izquierdo, A., Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology* 219(2), 607–20.

Izquierdo, A., Wellman, C. L., Holmes, A. (2006). Brief uncontrollable stress causes dendritic retraction in infralimbic cortex and resistance to fear extinction in mice. *J Neurosci* 26(21), 5733–8.

Izquierdo, A., Wiedholz, L. M., Millstein, R. A., Yang, R. J., Bussey, T. J., Saksida, L. M., et al. (2006). Genetic and dopaminergic modulation of reversal learning in a touchscreen-based operant procedure for mice. *Behav Brain Res* 171(2), 181–8.

Jordan, C. J., Harvey, R. C., Baskin, B. B., Dwoskin, L. P., Kantak, K. M. (2014). Cocaine-seeking behavior in a genetic model of attention-deficit/hyperactivity disorder following adolescent methylphenidate or Atomoxetine treatments. *Drug and Alcohol Dependence* 140, 25–32.

Keeler, J. F., Pretsell, D. O., Robbins, T. W. (2014). Functional implications of dopamine D1 vs. D2 receptors: A 'prepare and select' model of the striatal direct vs. indirect pathways. *Neuroscience* 282C, 156–175.

Kelley, A. E., Schochet, T., Landry, C. F. (2004). Risk taking and novelty seeking in adolescence: introduction to part I. *Ann N Y Acad Sci* 1021, 27–32.

Kosheleff, A. R., Rodriguez, D., O'Dell, S. J., Marshall, J. F., Izquierdo, A. (2012). Comparison of single-dose and extended methamphetamine administration on reversal learning in rats. *Psychopharmacology* 224(3), 459–467.

Larkin, J. D., Jenni, N. L., Floresco, S. B. (2015). Modulation of risk/reward decision making by dopaminergic transmission within the basolateral amygdala. *Psychopharmacology (Berl)* 233(1), 121-36

Laviola, G., Macrì, S., Morley-Fletcher, S., Adriani, W. (2003) Risk-taking behavior in adolescent mice: psychobiological determinants and early epigenetic influence. *Neurosci Biobehav Rev* 27(1–2), 19–31.

Marco, E. M., Adriani, W., Ruocco, L. A., Canese, R., Sadile, A. G., Laviola, G. (2011). Neurobehavioral adaptations to methylphenidate: the issue of early adolescent exposure. *Neurosci Biobehav Rev* 35(8), 1722–1739.

McDougall, S. A., Eaton, S. E., Mohd-Yusof, A., Crawford, C. A. (2015). Age-dependent changes in cocaine sensitivity across early ontogeny in male and female rats: possible role of dorsal striatal D2(High) receptors. *Psychopharmacology* 232(13), 2287–2301.

Mehta, M. A., Goodyer, I. M., Sahakian, B. J. (2004). Methylphenidate improves working memory and set-shifting in AD/HD: Relationships to baseline memory capacity. *J Child Psychol Psychiatry* 45(2), 293–305.

Miech, R. A., Johnston, L. D., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E. (2015). Monitoring the Future national survey results on drug use, 1975–2014: Volume I, Secondary school students. Ann Arbor: Institute for Social Research, The University of Michigan.

Muller, D., Mendez, P., Deroo, M., Klauser, P., Steen, S., and Poglia, L. (2010). Role of NCAM in spine dynamics and synaptogenesis. *Adv Exp Med Biol* 663, 245–256.

Muller, D., Wang, C., Skibo, G., Toni, N., Cremer, H., Calaora, V., et al. (1996). PSA-NCAM is required for activity-induced synaptic plasticity. *Neuron* 17, 413–422.

Nacher, J., Blasco-Ibáñez, J. M., and McEwen, B. S. (2002a). Non-granule PSA-NCAM immunoreactive neurons in the rat hippocampus. *Brain Res* 930, 1–11.

Nacher, J., Guirado, R., and Castillo-Gómez, E. (2013). Structural plasticity of interneurons in the adult brain: role of PSA-NCAM and implications for psychiatric disorders. *Neurochem Res* 38, 1122–1133.

Nacher, J., Lanuza, E., and McEwen, B. S. (2002b). Distribution of PSA-NCAM expression in the amygdala of the adult rat. *Neuroscience* 113, 479–484.

Norcross, M., Mathur, P., Enoch, A. J., Karlsson, R. M., Brigman, J. L., Cameron, H. A., et al. (2008). Effects of adolescent fluoxetine treatment on fear-, anxiety- or stress-related behaviors in C57BL/6J or BALB/cJ mice. *Psychopharmacology*. 200(3), 413–424.

Ochoa, J. G., Stolyarova, A., Kaur, A., Hart, E. E., Bugarin, A., Izquierdo, A. (2015). Post-training depletions of basolateral amygdala serotonin fail to disrupt discrimination, retention, or reversal learning. *Front Neurosci* 9, 155.

Ohno, Y., Okano, M., Masui, A., Imaki, J., Egawa, M., Yoshihara, C., et al. (2012). Region-specific elevation of D1 receptor-mediated neurotransmission in the nucleus accumbens of SHR, a rat model of attention deficit/hyperactivity disorder. *Neuropharmacology* 63(4), 547–554.

Pham, K., Nacher, J., Hof, P. R., and McEwen, B. S. (2003). Repeated restraint stress suppresses neurogenesis and induces biphasic PSA-NCAM expression in the adult rat dentate gyrus. *Eur J Neurosci* 17, 879–886.

Rygula, R., Clarke, H. F., Cardinal, R. N., Cockcroft, G. J., Xia, J., Dalley, J. W., et al. (2014). Role of central serotonin in anticipation of rewarding and punishing outcomes: Effects of selective amygdala or orbitofrontal 5-HT depletion. *Cereb Cortex* 25(9), 3064-3076.

Schaefer, T. L., Grace, C. E., Braun, A. A., Amos-Kroohs, R. M., Graham, D. L., Skelton, M. R., et al. (2013). Cognitive impairments from developmental exposure to serotonergic drugs: citalopram and MDMA. *Int J Neuropsychopharmacol* 16(6), 1383–1394.

Schweimer, J., Hauber, W. (2006) Dopamine D1 receptors in the anterior cingulate cortex regulate effort-based decision making. *Learn Mem* 13(6), 777–82.

Seki, T. (2002). Expression patterns of immature neuronal markers PSA-NCAM, CRMP-4 and NeuroD in the hippocampus of young adult and aged rodents. *J Neurosci Res* 70, 327–334. doi: 10.1002/jnr.10387

Shanks, R. A., Ross, J. M., Doyle, H. H., Helton, A. K., Picou, B. N., Schulz, J., et al. (2015). Adolescent exposure to cocaine, amphetamine, and methylphenidate cross-sensitizes adults to methamphetamine with drug- and sex-specific effects. *Behavioural Brain Research* 281, 116–124.

Spear, L. P. (2015). Adolescent alcohol exposure: Are there separable vulnerable periods within adolescence? *Physiol Behav* 148, 122–130.

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neurosci Biobehav Rev* 24(4), 417–463.

Stolyarova, A., O'Dell, S. J., Marshall, J. F., Izquierdo, A. (2014). Positive and negative feedback learning and associated dopamine and serotonin transporter binding after methamphetamine. *Behav Brain Res.* 271, 195–202.

Stolyarova, A., Izquierdo A. (2015). Distinct patterns of outcome valuation and amygdala-prefrontal cortex synaptic remodeling in adolescence and adulthood. *Front Behav Neurosci* 9, 115.

Stopper, C. M., Khayambashi, S., Floresco, S. B. (2013). Receptor-specific modulation of risk-based decision making by nucleus accumbens dopamine. *Neuropsychopharmacology* 38(5), 715–28.

Tarazi, F. I., Baldessarini, R. J. (2000). Comparative postnatal development of dopamine D(1), D(2) and D(4) receptors in rat forebrain. *Int J Dev Neurosci* 18(1), 29–37.

Teicher, M. H., Andersen, S. L., Hostetter, J. C., Jr. (1995). Evidence for dopamine receptor pruning between adolescence and adulthood in striatum but not nucleus accumbens. *Brain Res Dev Brain Res* 89, 167–172.

Thanos, P. K., Michaelides, M., Benveniste, H., Wang, G. J., Volkow, N. D. (2007). Effects of chronic oral methylphenidate on cocaine self-administration and striatal dopamine D2 receptors in rodents. *Pharmacol Biochem* Behav 87(4), 426–433.

Tye, K. M., Tye, L. D., Cone, J. J., Hekkelman, E. F., Janak, P. H., Bonci, A. (2010) Methylphenidate facilitates learning-induced amygdala plasticity. *Nat Neurosci* 13(4), 475–81.

Varea, E., Nácher, J., Blasco-Ibáñez, J. M., Gómez-Climent, M. A., Castillo-Gómez, E., Crespo, C., et al. (2005). PSA-NCAM expression in the rat medial prefrontal cortex. *Neuroscience* 136, 435–443.

Vendruscolo, L. F., Izídio, G. S., Takahashi, R. N., Ramos, A. (2008). Chronic methylphenidate treatment during adolescence increases anxiety-related behaviors and ethanol drinking in adult spontaneously hypertensive rats. *Behavioural Pharmacology* 19(1), 21–27.

Vorhees, C. V., Morford, L. R., Graham, D. L., Skelton, M. R., Williams, M. T. (2011). Effects of periadolescent fluoxetine and paroxetine on elevated plus-maze, acoustic startle, and swimming immobility in rats while on and off-drug. *Behavioral and Brain Functions* 7, 41.

Vutskits, L., Djebbara-Hannas, Z., Zhang, H., Paccaud, J. P., Durbec, P., Rougon, G., et al. (2001). PSA-NCAM modulates BDNF-dependent survival and differentiation of cortical neurons. *Eur J Neurosci* 13, 1391–1402.

Wass, C., Pizzo, A., Sauce, B., Kawasumi, Y., Sturzoiu, T., Ree, F., et al. (2013). Dopamine D1 sensitivity in the prefrontal cortex predicts general cognitive abilities and is modulated by working memory training. *Learn Mem* 20(11), 617–627.

Wassum, K. M., Izquierdo, A. (2015). The basolateral amygdala in reward learning and addiction. *Neurosci Biobehav Rev* 57, 271-283.

Xiao, M. F., Xu, J. C., Tereshchenko, Y., Novak, D., Schachner, M., and Kleene, R. (2009). Neural cell adhesion molecule modulates dopaminergic signaling and behavior by regulating dopamine D2 receptor internalization. *J Neurosci* 29, 14752–14763.

Ye, T., Pozos, H., Phillips, T. J., Izquierdo, A. (2014). Long-term effects of exposure to methamphetamine in adolescent rats. *Drug Alcohol Depend* 138, 17–23.

Yohn, S. E., Santerre, J. L., Nunes, E. J., Kozak, R., Podurgiel, S. J., Correa, M., et al. (2015). The role of dopamine D1 receptor transmission in effort-related choice behavior: Effects of D1 agonists. *Pharmacol Biochem Behav* 135, 217–26.

Zito, J. M., Safer, D. J., dosReis, S., Gardner, J. F., Boles, M., Lynch, F. (2000). Trends in the prescribing of psychotropic medications to preschoolers. *JAMA* 283(8), 1025–1030.

Chapter 4: The credit assignment problem

Abstract

In naturalistic multi-cue and multi-step learning tasks, where outcomes of behavior are delayed in time, discovering which choices are responsible for rewards can present a challenge, known as the *credit assignment problem*. In this review, I summarize recent work that highlighted a critical role for the prefrontal cortex (PFC) in assigning credit where it is due in tasks where only a few of the multitude of cues or choices are relevant to the final outcome of behavior. Collectively, these investigations have provided compelling support for specialized roles of the orbitofrontal (OFC), anterior cingulate (ACC), and dorsolateral prefrontal (dlPFC) cortices in contingent learning. However, recent work has similarly revealed shared contributions and emphasized rich and heterogeneous response properties of neurons in these brain regions. Such functional overlap is not surprising given the complexity of reciprocal projections spanning the PFC. In the concluding section, I overview the evidence suggesting that the OFC, ACC and dlPFC communicate extensively, sharing information about presented options, executed decisions and received rewards, which enables them to assign credit for outcomes to choices on which they are contingent. This account suggests that lesion or inactivation/inhibition experiments targeting a localized PFC subregion will be insufficient to gain a fine-grained understanding of credit assignment during learning and instead poses refined questions for future research, shifting the focus from focal manipulations to experimental techniques targeting cortico-cortical projections.

Introduction

When an animal is introduced to an unfamiliar environment, it will explore the surroundings randomly until an unexpected reward is encountered. Reinforced by this experience, the animal will gradually learn to repeat those actions that produced the desired outcome. The work conducted in the past several decades has contributed to a detailed understanding of the psychological and neural mechanisms that support such reinforcement-driven learning (Schultz and Dickinson, 2000; Schultz, 2004; Niv, 2009). It is now broadly accepted that dopamine (DA) signaling conveys prediction errors, or the degree of surprise brought about by unexpected rewards, and interacts with cortical and basal ganglia circuits to selectively reinforce the advantageous choices (Schultz, 1998a,b; Schultz and Dickinson, 2000; Niv, 2009). Yet, in naturalistic settings, where rewards are delayed in time, multiple cues are encountered, or several decisions are made before the outcomes of behavior are revealed, discovering which choices are responsible for rewards can present a challenge, known as the *credit assignment problem* (Mackintosh, 1975; Rothkopf and Ballard, 2010).

In most everyday situations, rewards are not immediate consequences of behavior, but instead appear after substantial delays. To influence future choices, the teaching signal conveyed by DA release needs to reinforce synaptic events occurring on a millisecond timescale, frequently seconds before the outcomes of decisions are revealed (Izhikevich, 2007; Fisher et al., 2017). This apparent difficulty in linking preceding behaviors caused by transient neuronal activity to delayed feedback has been termed the *distal reward* or *temporal credit assignment problem* (Hull, 1943; Barto et al., 1983; Sutton and Barto, 1998; Dayan and Abbott, 2001; Wörgötter and Porr, 2005). Credit for the reward delayed by several seconds can frequently be assigned by establishing an eligibility trace, a molecular memory of the recent neuronal activity, allowing modification of

synaptic connections that participated in the behavior (Pan et al., 2005; Fisher et al., 2017). On longer timescales, or when multiple actions need to be performed sequentially to reach a final goal, intermediate steps themselves can acquire motivational significance and subsequently reinforce preceding decisions, such as in temporal-difference (TD) learning models (Sutton and Barto, 1998).

Several excellent reviews have summarized the accumulated knowledge on mechanisms that link choices and their outcomes through time, highlighting the advantages of eligibility traces and TD models (Wörgötter and Porr, 2005; Barto, 2007; Niv, 2009; Walsh and Anderson, 2014). Yet these solutions to the distal reward problem can impede learning in multi-choice tasks, or when an animal is presented with many irrelevant stimuli prior to or during the delay. Here, I only briefly overview the work on the distal reward problem to highlight potential complications that can arise in credit assignment based on eligibility traces when learning in multi-cue environments. Instead, I focus on the *structural* (or *spatial*) credit assignment problem, requiring animals to select and learn about the most meaningful features in the environment and ignore irrelevant distractors. Collectively, the reviewed evidence highlights a critical role for the prefrontal cortex (PFC) in such contingent learning.

Recent studies have provided compelling support for specialized functions of the orbitofrontal (OFC) and dorsolateral prefrontal (dlPFC) cortices in credit assignment in multi-cue tasks, with fewer experiments targeting the anterior cingulate cortex (ACC). For example, it has been suggested that the dlPFC aids reinforcement-driven learning by directing attention to task-relevant cues (Niv et al., 2015), the OFC assigns credit for rewards based on the causal relationship between trial outcomes and choices (Jocham et al., 2016; Noonan et al., 2017), and the ACC contributes to unlearning of action-outcome associations when the rewards are available for free

(Jackson et al., 2016). However, this work has similarly revealed shared contributions and emphasized rich and heterogeneous response properties of neurons in the PFC, with different subregions monitoring and integrating the information about the task (i.e., current context, available options, anticipated rewards, delay and effort costs) at variable times within a trial (upon stimulus presentation, action selection, outcome anticipation, and feedback monitoring; ex., Hunt et al., 2015; Khamassi et al., 2015). In the concluding section, I overview evidence suggesting that contingent learning in multi-cue environments relies on dynamic cortico-cortical interactions during decision making and outcome valuation.

Solving the temporal credit assignment problem

When outcomes follow choices after short delays (**Figure 4-1A**), the credit for distal rewards can frequently be assigned by establishing an eligibility trace, a sustained memory of the recent activity that renders synaptic connections malleable to modification over several seconds. Eligibility traces can persist as elevated levels of calcium in dendritic spines of post-synaptic neurons (Kötter and Wickens, 1995) or as sustained neuronal activity throughout the delay period (Curtis and Lee, 2010) to allow for synaptic changes in response to reward signals. Furthermore, spike-timing dependent plasticity can be influenced by neuromodulator input (Izhikevich, 2007; Abraham, 2008; Fisher et al., 2017). For example, the magnitude of short-term plasticity can be modulated by DA, acetylcholine and noradrenaline, which may even revert the sign of the synaptic change (Matsuda et al., 2006; Izhikevich, 2007; Seol et al., 2007; Abraham, 2008; Zhang et al., 2009). Sustained neural activity has been observed in the PFC and striatum (Jog et al., 1999; Pasupathy and Miller, 2005; Histed et al., 2009; Kim et al., 2009, 2013; Seo et al., 2012; Her et al., 2016), as well as the sensory cortices after experience with consistent pairings between stimuli and outcomes separated by predictable delays (Shuler and Bear, 2006).

On extended timescales, when multiple actions need to be performed sequentially to reach a final goal, the distal reward problem can be solved by assigning motivational significance to intermediate choices that can subsequently reinforce preceding decisions, such as in TD learning models (Montague et al., 1996; Sutton and Barto, 1998; Barto, 2007). Assigning values to these intervening steps according to expected future rewards allows the learner to break complex temporal credit assignment problems into smaller and easier tasks. There is ample evidence for TD learning in humans and other animals that on the neural level is supported by transfer of DA

responses from the time of reward delivery to preceding cues and actions (Montague et al., 1996; Schultz, 1998a,b; Walsh and Anderson, 2014).

Both TD learning and eligibility traces offer elegant solutions to the distal reward problem, and models based on cooperation between these two mechanisms can predict animal behavior as well as neuronal responses to rewards and predictive stimuli (Pan et al., 2005; Bogacz et al., 2007). Yet assigning credit based on eligibility traces can be suboptimal when an animal interacts with many irrelevant stimuli prior to or during the delay (**Figure 4-1B**). Under such conditions sensory areas remain responsive to distracting stimuli and the arrival of non-specific reward signals can reinforce intervening cues that did not meaningfully contribute to, but occurred close to, the outcome of behavior (FitzGerald et al., 2013; Xu, 2017).
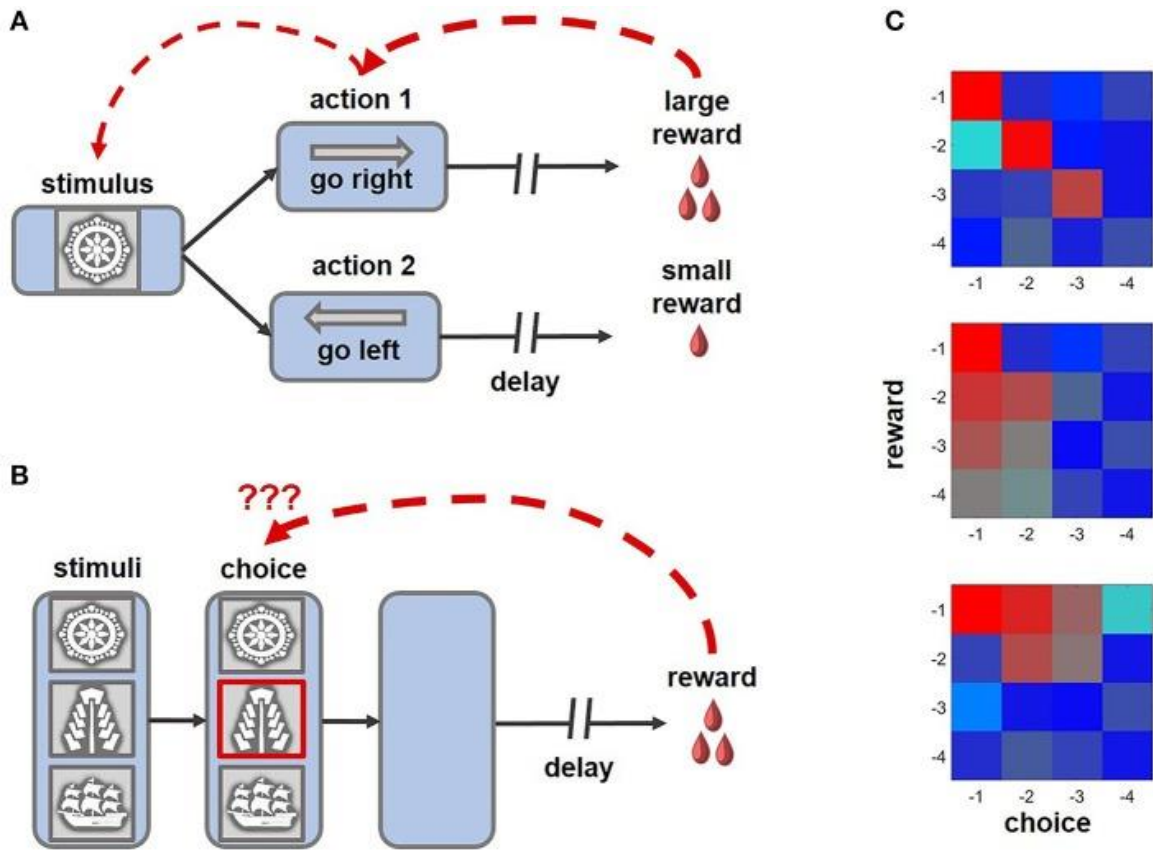
Figure 4-1. Example tasks highlighting the challenge of credit assignment and learning strategies enabling animals to solve this problem. (A) An example of a distal reward task that can be successfully learned with eligibility traces and TD rules, where intermediate choices can acquire motivational significance and subsequently reinforce preceding decisions (ex., Pasupathy and Miller, 2005; Histed et al., 2009). **(B)** In this version of the task, multiple cues are present at the time of choice, only one of which is meaningful for obtaining rewards. After a brief presentation, the stimuli disappear, requiring an animal to solve a complex structural and temporal credit assignment problem (ex., Noonan et al., 2010, 2017; Niv et al., 2015; Asaad et al., 2017; while the schematic of the task captures the challenge of credit assignment, note that in some experimental variants of the behavioral paradigm stimuli disappeared before an animal revealed its choice, whereas in others the cues remained on the screen until the trial outcome was revealed). Under such conditions, learning based on eligibility traces is suboptimal, as non-specific reward signals can reinforce visual cues that did not meaningfully contribute, but occurred close, to beneficial outcomes of behavior. **(C)** On reward tasks, similar to the one shown in **(B)**, the impact of previous decisions and associated rewards on current behavior can be assessed by performing regression analyses (Jocham et al., 2016; Noonan et al., 2017). Here, the color of each cell in a matrix represents the magnitude of the effect of short-term choice and outcome histories, up to 4 trials into the past (red-strong influence; blue-weak influence on the current decision). **Top**: an animal learning based on the causal relationship between outcomes and choices (i.e., contingent learning). **Middle**: each choice is reinforced by a combined history of rewards (i.e., decisions are repeated if beneficial outcomes occur frequently). **Bottom**: the influence of recent rewards spreads to unrelated choices.

The role of the PFC in structural credit assignment

Several recent studies have investigated the neural mechanisms of appropriate credit assignment in challenging tasks where only a few of the multitude of cues predict rewards reliably. Collectively, this work has provided compelling support for causal contributions of the PFC to structural credit assignment. For example, Asaad et al. (2017) examined the activity of neurons in monkey dlPFC while subjects were performing a delayed learning task. The arrangement of the stimuli varied randomly between trials and within each block either the spatial location or stimulus identity was relevant for solving the task. The monkeys' goal was to learn by trial-and-error to select one of the four options that led to rewards according to current rules. When stimulus identity was relevant for solving the task, neural activity in the dlPFC at the time of feedback reflected both the relevant cue (regardless of its spatial location) and the trial outcome, thus integrating the information necessary for credit assignment. Such responses were strategy-selective: these neurons did not encode cue identity at the time of feedback when it was not necessary for learning in the spatial location task, in which making a saccade to the same position on the screen was reinforced within a block of trials. Previous research has similarly indicated that neurons in the dlPFC respond selectively to behaviorally-relevant and attended stimuli (Lebedev et al., 2004; Markowitz et al., 2015) and integrate information about prediction errors, choice values as well as outcome uncertainty prior to trial feedback (Khamassi et al., 2015).

Activity within the dlPFC has been linked to structural credit assignment through selective attention and representational learning (Niv et al., 2015). Under conditions of reward uncertainty and unknown relevant task features, human participants opt for computational efficiency and engage in a serial-hypothesis-testing strategy (Wilson and Niv, 2011), selecting one cue and its anticipated outcome as the main focus of their behavior and updating the expectations associated

exclusively with that choice upon feedback receipt (Akaishi et al., 2016). Niv and colleagues tested participant on a three-armed bandit task, where relevant stimulus dimensions (i.e., shape, color or texture) predicting outcome probabilities changed between block of trials (Niv et al., 2015). In such a multidimensional environment, reinforcement-driven learning was aided by attentional control mechanisms that engaged the dlPFC, intraparietal cortex, and precuneus.

In many tasks, credit for outcomes can be assigned according to different rules: based on the causal relationship between rewards and choices (i.e., contingent learning), their temporal proximity (i.e., when the reward is received shortly after a response), or their statistical relationship (when an action has been executed frequently before beneficial outcomes; Jocham et al., 2016; **Figure 4-1C**). The analyses presented in the papers discussed above did not allow for dissociation between these alternative strategies of credit assignment. By testing human participants on a task with continuous stimulus presentation, instead of a typical trial-by-trial structure, Jocham et al. (2016) demonstrated that the tendency to repeat choices that were immediately followed by rewards and causal learning operate in parallel. In this experiment, activity within another subregion of the PFC, the OFC, was associated with contingent learning. Complementary work in monkeys revealed that the OFC contributes causally to credit assignment (Noonan et al., 2010): animals with OFC lesions were unable to associate a reward with the choice on which it was contingent and instead relied on temporal and statistical learning rules. In another recent paper, Noonan and colleagues (2017) extended these observations to humans, demonstrating causal contributions of the OFC to credit assignment across species. The participants were tested on a three-choice probabilistic learning task. The three options were presented simultaneously and maintained on the screen until the outcome of a decision was revealed, thus requiring participants to ignore irrelevant distractors. Notably, only patients with lateral OFC lesions displayed any

difficulty in learning the task, whereas damage to the medial OFC or dorsomedial PFC preserved contingent learning mechanisms. However, it is presently unknown whether lesions to the dlPFC or ACC affect such causal learning.

In another test of credit assignment in learning, contingency degradation, the subjects are required to track causal relationships between stimuli or actions and rewards. During contingency degradation sessions, the animals are still reinforced for responses, but rewards are also available for free. After experiencing non-contingent rewards, control subjects reliably decrease their choices of the stimuli. However, lesions to both the ACC and OFC inhibit contingency degradation (Jackson et al., 2016). Taken together, these observations demonstrate causal contributions of the PFC to appropriate credit assignment in multi-cue environments.

Cooperation between PFC subregions supports contingent learning in multi-cue tasks

Despite the segregation of temporal and structural aspects of credit assignment in earlier sections of this review, in naturalistic settings brains frequently need to tackle both problems simultaneously. Here, I overview the evidence favoring a network perspective, suggesting that dynamic cortico-cortical interactions during decision making and outcome valuation enable adaptive solutions to complex spatio-temporal credit assignment problems. It has been previously suggested that feedback projections from cortical areas occupying higher levels of the processing hierarchy, including the PFC, can aid in attribution of outcomes to individual decisions by implementing attention-gated reinforcement learning (Roelfsema and van Ooyen, 2005). Similarly, recent theoretical work has shown that even complex multi-cue and multi-step problems can be solved by an extended cascade model of synaptic memory traces, in which the plasticity is modulated not only by activity within a population of neurons but also by feedback about executed decisions and resulting rewards (Urbanczik and Senn, 2009; Friedrich et al., 2010, 2011). Contingent learning, according to these models, can be supported by the communication between neurons encoding available options, committed choices and outcomes of behavior during decision making and feedback monitoring. For example, at the time of outcome valuation, information about recent choices can be maintained as a memory trace in the neuronal population involved in action selection or conveyed by an efference copy from an interconnected brain region (Curtis and Lee, 2010; Khamassi et al., 2011, 2015). Similarly, reinforcement feedback is likely communicated as a global reward signal (ex., DA release) as well as projections from neural populations engaged in performance monitoring, such as those within the ACC (Friedrich et al., 2010; Khamassi et al., 2011). The complexity of reciprocal and recurrent projections spanning the PFC (Barbas and Pandya, 1989; Felleman and Van Essen, 1991; Elston, 2000) may enable this

network to implement such learning rules, integrating information about the task, executed decisions and performance feedback.

In many everyday decisions, options are compared across multiple features simultaneously (ex., by considering current context, needs, available reward types, and delay and effort costs). Neurons in different subregions of the PFC exhibit rich response properties, signaling these features of the task at various time epochs within a trial. For example, reward selectivity in response to predictive stimuli emerges earlier in the OFC and may then be passed to the dlPFC which encodes both the expected outcome and the upcoming choice (Wallis and Miller, 2003). Similarly, on trials where options are compared based on delays to rewards, choices are dependent on interactions between the OFC and dlPFC (Hunt et al., 2015). Conversely, when effort costs are more meaningful for decisions, it is the ACC that influences choice-related activity in the dlPFC (Hunt et al., 2015). The OFC is required not only for the evaluation of stimuli, but also more complex abstract rules, based on rewards they predict (Buckley et al., 2009). While both the OFC and dlPFC encode abstract strategies (ex., persisting with recent choices or shifting to a new response), such signals appear earlier in the OFC and may be subsequently conveyed to the dlPFC where they are combined with upcoming response (i.e., left vs. right saccade) encoding (Tsujimoto et al., 2011). Therefore, the OFC may be the first PFC subregion to encode task rules and/or potential rewards predicted by sensory cues; via cortico-cortical projections, this information may be subsequently communicated to the dlPFC or ACC (Kennerley et al., 2009; Hayden and Platt, 2010) to drive strategy-sensitive response planning.

The behavioral strategy that the animal follows is influenced by recent reward history (Cohen et al., 2007; Pearson et al., 2009). If its choices are reinforced frequently, the animal will make similar decisions in the future (i.e., exploit its current knowledge). Conversely, unexpected

omission of expected rewards can signal a need for novel behaviors (i.e., exploration). Neurons in the dlPFC carry representations of planned as well as previous choices, anticipate outcomes, and jointly encode the current decisions and their consequences following feedback (Seo and Lee, 2007; Seo et al., 2007; Tsujimoto et al., 2009; Asaad et al., 2017). Similarly, the ACC tracks trial-by-trial outcomes of decisions (Procyk et al., 2000; Shidara and Richmond, 2002; Amiez et al., 2006; Quilodran et al., 2008) as well as reward and choice history (Seo and Lee, 2007; Kennerley et al., 2009, 2011; Sul et al., 2010; Kawai et al., 2015) and signals errors in outcome prediction (Kennerley et al., 2009, 2011; Hayden et al., 2011; Monosov, 2017). At the time of feedback, neurons in the OFC encode committed choices, their values and contingent rewards (Tsujimoto et al., 2009; Sul et al., 2010). Notably, while the OFC encodes the identity of expected outcomes and the value of the chosen option after the alternatives are presented to an animal, it does not appear to encode upcoming decisions (Tremblay and Schultz, 1999; Wallis and Miller, 2003; Padoa-Schioppa and Assad, 2006; Sul et al., 2010; McDannald et al., 2014), therefore it might be that feedback projections from the dlPFC or ACC are required for such activity to emerge at the time of reward feedback.

To capture the interactions between PFC subregions in reinforcement-driven learning, Khamassi and colleagues have formulated a computational model in which action values are stored and updated in the ACC and then communicated to the dlPFC which decides which action to trigger (Khamassi et al., 2011, 2013). This model relies on meta-learning principles (Doya, 2002), flexibly adjusting the exploration-exploitation parameter based on performance history and variability in the environment which are monitored by the ACC. The explore-exploit parameter then influences action-selection mechanisms in the dlPFC, prioritizing choice repetition once the rewarded actions are discovered and encouraging switching between different options when

environmental conditions change. In addition to highlighting the dynamic interactions between the dlPFC and ACC in learning, the model similarly offers an elegant solution to the credit assignment problem by restricting value updating only to those actions that were selected on a given trial. This is implemented by requiring the prediction error signals in the ACC to coincide with a motor efference copy sent by the premotor cortex. The model is endorsed with an ability to learn meta-values of novel objects in the environment based on changes in the average reward that follow the presentation of such stimuli. While the authors proposed that such meta-value learning is implemented by the ACC, it is plausible that the OFC also plays a role in this process based on its contributions to stimulus-outcome and state learning (Wilson et al., 2014; Zsuga et al., 2016). Intriguingly, this model could reproduce monkey behavior and neural responses on two tasks: four-choice deterministic and two-choice probabilistic paradigms, entailing a complex spatio-temporal credit assignment problem as the stimuli disappeared from the screen prior to action execution and outcome presentation (Khamassi et al., 2011, 2013, 2015). Model-based analyses of neuronal responses further revealed that information about prediction errors, action values and outcome uncertainty is integrated both in the dlPFC and ACC, but at different timepoints: before trial feedback in the dlPFC and after feedback in the ACC (Khamassi et al., 2015).

Collectively, these findings highlight the heterogeneity of responses in each PFC subregion that differ in temporal dynamics within a single trial and suggest that the cooperation between the OFC, ACC and dlPFC may support flexible strategy- and context-dependent choices. This network perspective further suggests that individual PFC subregions may be less specialized in their functions than previously thought. For example, in primates both the ACC and dlPFC participate in decisions based on action values (Hunt et al., 2015; Khamassi et al., 2015). More recently, it has been demonstrated that the OFC is involved in updating action-outcome values as well (Fiuzat

et al., 2017). Analogously, while it has been proposed that the OFC is specialized for stimulus-outcome and ACC for action-outcome learning (Rudebeck et al., 2008), lesions to the ACC have been similarly reported to impair stimulus-based reversal learning (Chudasama et al., 2013), supporting shared contributions of the PFC subregions to adaptive behavior. Indeed, these brain regions communicate extensively, sharing information about presented options, executed decisions and received rewards (**Figure 4-2**), which can enable them to assign credit for outcomes to choices on which they are contingent (Urbanczik and Senn, 2009; Friedrich et al., 2010, 2011). Attention-gated learning likely relies on cooperation between PFC subregions as well: for example, coordinated and synchronized activity between the ACC and dlPFC aids in goal-directed attentional shifting and prioritization of task-relevant information (Womelsdorf et al., 2014; Oemisch et al., 2015; Voloh et al., 2015).

Functional connectivity within the PFC can support contingent learning on shorter timescales (ex., across trials within the same task), when complex rules or stimulus-action-outcome mappings are switching frequently (Duff et al., 2011; Johnson et al., 2016). Under such conditions, the same stimuli can carry different meaning depending on task context or due to changes in the environment (ex., serial discrimination-reversal problems) and PFC neurons with heterogeneous response properties may be better targets for modification, allowing the brain to exert flexible, rapid and context-sensitive control over behavior (Asaad et al., 1998; Mansouri et al., 2006). Indeed, it has been shown that rule and reversal learning induce plasticity in OFC synapses onto the dorsomedial PFC (encompassing the ACC) in rats (Johnson et al., 2016). When motivational significance of reward-predicting cues fluctuates frequently, neuronal responses and synaptic connections within the PFC tend to update more rapidly (i.e., across block of trials) compared to subcortical structures and other cortical regions (Padoa-Schioppa and Assad, 2008;

Morrison et al., 2011; Xie and Padoa-Schioppa, 2016; Fernández-Lamo et al., 2017; Saez et al., 2017). Similarly, neurons in the PFC promptly adapt their responses to incoming information based on the recent history of inputs (Freedman et al., 2001; Meyers et al., 2012; Stokes et al., 2013). Critically, changes in the PFC activity closely track behavioral performance (Mulder et al., 2003; Durstewitz et al., 2010), and interfering with neural plasticity within this brain area prevents normal responses to contingency degradation (Swanson et al., 2015).

When circumstances are stable overall and the same cues or actions remain reliable predictors of rewards, long-range connections between the PFC, association and sensory areas can support contingent learning on prolonged timescales. Neurons in the lateral intraparietal area demonstrate larger post-decisional responses and enhanced learning following choices that predict final outcomes of sequential behavior in a multi-step and -cue task (Gersch et al., 2014). Such changes in neuronal activity likely rely on information about task rules conveyed by the PFC directly or via interactions with neuromodulatory systems. These hypotheses could be tested in future work.

In summary, dynamic interactions between subregions of the PFC can support contingent learning in multi-cue environments. Furthermore, via feedback projections, the PFC can guide plasticity in other cortical areas associated with sensory and motor processing (Cohen et al., 2011). This account suggests that lesion experiments targeting a localized PFC subregion will be insufficient to gain fine-grained understanding of credit assignment during learning and instead poses refined questions for future research, shifting the focus from focal manipulations to experimental techniques targeting cortico-cortical projections. To gain novel insights into functional connectivity between PFC subregions, it will be critical to assess neural correlates of contingent learning in the OFC, ACC, and dlPFC simultaneously in the context of the same task.

upcoming choice,
expected outcome,
abstract rule-strategy,
executed decision,
choice-reward history

dlPFC

ACC

OFC

upcoming choice,
effort cost,
choice-reward history,
received outcome,
reward prediction error

expected outcome,
delay cost,
abstract rule-strategy,
executed decision,
received outcome
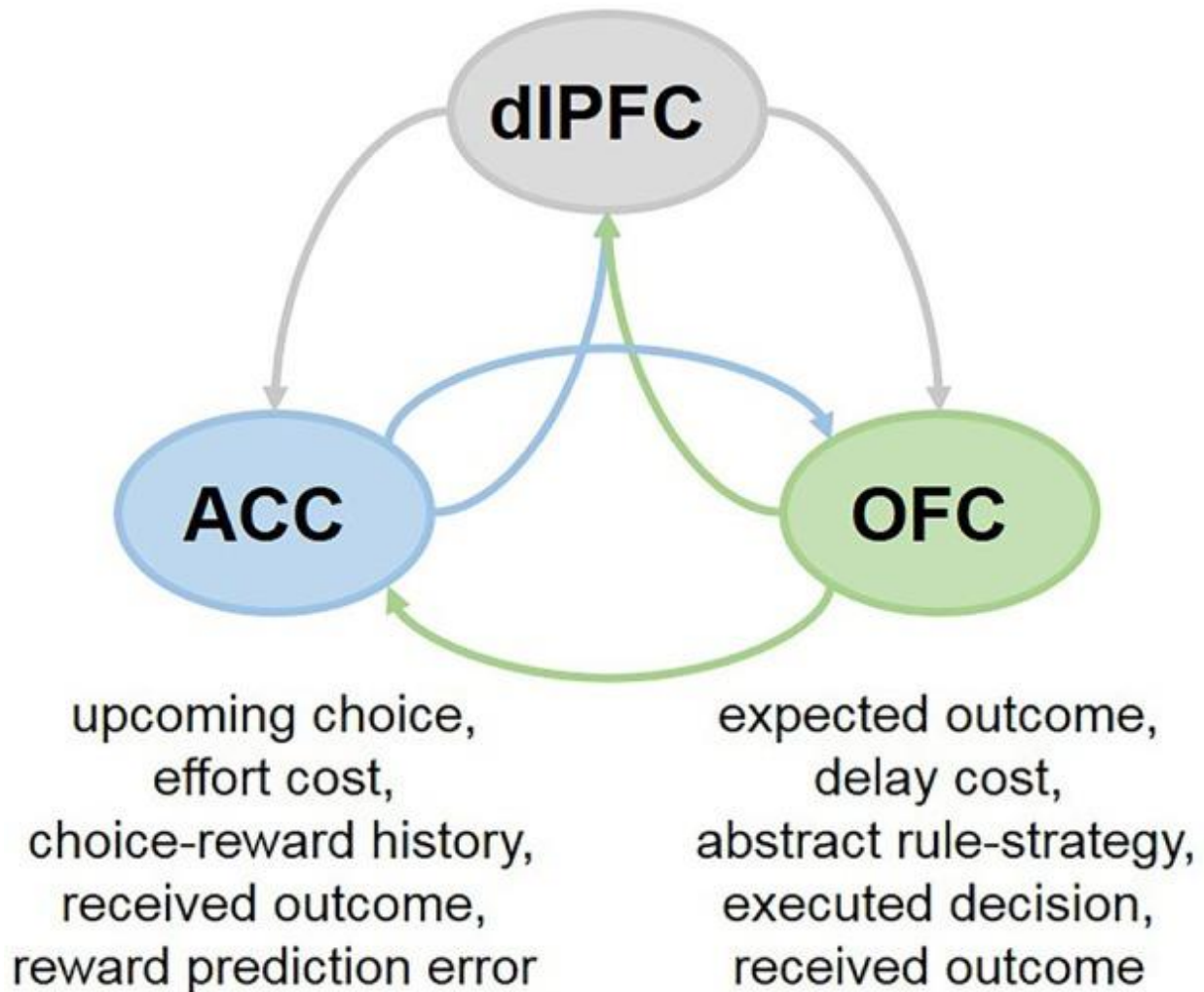
Figure 4-2. Cooperation between PFC subregions in multi-cue tasks. In many everyday decisions, the options are compared across multiple features simultaneously (ex., by considering current context, needs, available reward types, as well as delay and effort costs). Neurons in different subregions of the PFC exhibit rich response properties, integrating many aspects of the task at hand. The OFC, ACC and dlPFC communicate extensively, sharing the information about presented options, executed decisions and received rewards, which can enable them to assign credit for outcomes to choices on which they are contingent.

In humans, functional connectivity can be assessed by utilizing coherence, phase synchronization, Granger causality and Bayes network approaches (Bastos and Schoffelen, 2016; Mill et al., 2017). Indeed, previous studies have linked individual differences in cortico-striatal functional connectivity to reinforcement-driven learning (Horga et al., 2015; Kaiser et al., 2017) and future work could focus on examining cortico-cortical interactions in similar paradigms. To probe causal contributions of projections spanning the PFC, future research may benefit from designing multi-cue tasks for rodents and taking advantage of recently developed techniques (i.e., chemo- and opto-genetic targeting of projection neurons followed by silencing of axonal terminals to achieve pathway-specific inhibition; Deisseroth, 2010; Sternson and Roth, 2014) that afford increasingly precise manipulations of cortico-cortical connectivity. It should be noted, however, that most experiments to date have probed the contributions of the PFC to credit assignment in primates, and functional specialization across different subregions may be even less pronounced in mice and rats. Finally, as highlighted throughout this review, the recent progress in understanding the neural mechanisms of credit assignment has relied on the introduction of more complex tasks, including multi-cue and probabilistic choice paradigms. While such tasks better mimic the naturalistic problems that the brains have evolved to solve, they also produce behavioral patterns that are more difficult to analyze and interpret (Scholl and Klein-Flügge, 2017). As such, computational modeling of behavior and neuronal activity may prove especially useful in future work on credit assignment.

References

Abraham, W. C. (2008). Metaplasticity: tuning synapses and networks for plasticity. *Nat Rev Neurosci* 9, 387.

Akaishi, R., Kolling, N., Brown, J. W., Rushworth, M. (2016). Neural mechanisms of credit assignment in a multicue environment. *J Neurosci* 36, 1096–1112.

Amiez, C., Joseph, J. P., Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex* 16, 1040–1055.

Asaad, W. F., Lauro, P. M., Perge, J. A., Eskandar, E. N. (2017). Prefrontal neurons encode a solution to the credit assignment problem. *J Neurosci* 37, 6995–7007.

Asaad, W. F., Rainer, G., Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21, 1399–1407.

Barbas, H., Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 286, 353–375.

Barto, A. G. (2007). Temporal difference learning. *Scholarpedia* J 2, 1604.

Barto, A. G., Sutton, R. S., Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems, in *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 834–846.

Bastos, A. M., Schoffelen, J. M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front Syst Neurosci* 9, 175.

Bogacz, R., McClure, S. M., Li, J., Cohen, J. D., Montague, P. R. (2007). Short-term memory traces for action bias in human reinforcement learning. *Brain Res* 1153, 111–121.

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G. F., Kwok, S. C., et al. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325, 52–58.

Chudasama, Y., Daniels, T. E., Gorrin, D. P., Rhodes, S. E., Rudebeck, P. H., Murray, E. A. (2013). The role of the anterior cingulate cortex in choices based on reward value and reward contingency. *Cereb Cortex* 23, 2884–2898.

Cohen, J. D., McClure, S. M., Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc Lond Ser B Biol Sci* 362, 933–942.

Cohen, M. X., Wilmes, K., van de Vijver, I. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends Cogn Sci* 15, 558–566.

Curtis, C. E., Lee, D. (2010). Beyond working memory: the role of persistent activity in decision making. *Trends Cogn Sci* 14, 216–222.

Dayan, P., Abbott, L. F. (2001). Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. Cambridge, MA: MIT Press.

Deisseroth, K. (2010). Optogenetics. *Nat Methods* 8, 26–29.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Netw* 15, 495–506.

Duff, A., Sanchez, Fibla M., Verschure, P. F. M. J. (2011). A biologically based model for the integration of sensory–motor contingencies in rules and plans: a prefrontal cortex based extension of the distributed adaptive control architecture. *Brain Res Bull* 85, 289–304.

Durstewitz, D., Vittoz, N. M., Floresco, S. B., Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66, 438–448.

Elston, G. N. (2000). Pyramidal cells of the frontal lobe: all the more spinous to think with. *J Neurosci* 20, RC95.

Felleman, D. J., Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1, 1–47.

Fernández-Lamo, I., Delgado-García, J. M., Gruart, A. (2017). When and where learning is taking place: multisynaptic changes in strength during different behaviors related to the acquisition of an operant conditioning task by behaving rats. *Cereb Cortex* 14, 1–13.

Fisher, S. D., Robertson, P. B., Black, M. J., Redgrave, P., Sagar, M. A., Abraham, W. C., *et al.* (2017). Reinforcement determines the timing dependence of corticostriatal synaptic plasticity in vivo. *Nat Commun* 8, 334.

FitzGerald, T. H. B., Friston, K. J., Dolan, R. J. (2013). Characterising reward outcome signals in sensory cortex. *NeuroImage* 83, 329–334.

Fiuzat, E. C., Rhodes, S. E., Murray, E. A. (2017). The role of orbitofrontal-amygdala interactions in updating action-outcome valuations in macaques. *J Neurosci* 37, 2463–2470.

Freedman, D. J., Riesenhuber, M., Poggio, T., Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316.

Friedrich, J., Urbanczik, R., Senn, W. (2010). Learning spike-based population codes by reward and population feedback. *Neural Comput* 22, 1698–1717.

Friedrich, J., Urbanczik, R., Senn, W. (2011). Spatio-temporal credit assignment in neuronal population learning. *PLoS Comput Biol* 7, e1002092.

Gersch, T. M., Foley, N. C., Eisenberg, I., Gottlieb, J. (2014). Neural correlates of temporal credit assignment in the parietal lobe. *PloS One* 9, e88725.

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., Platt, M. L. (2011). Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *J Neurosci* 31, 4178–4187.

Hayden, B. Y., Platt, M. L. (2010). Neurons in anterior cingulate cortex multiplex information about reward and action. *J Neurosci* 30, 3339–3346.

Her, E. S., Huh, N., Kim, J., Jung, M. W. (2016). Neuronal activity in dorsomedial and dorsolateral striatum under the requirement for temporal credit assignment. *Sci. Rep.* 6, 27056.

Histed, M. H., Pasupathy, A., Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63, 244–253.

Horga, G., Maia, T. V., Marsh, R., Hao, X., Xu, D., Duan, Y., et al. (2015). Changes in corticostriatal connectivity during reinforcement learning in humans. *Hum Brain Mapp* 36, 793–803.

Hull, C. (1943). Principles of Behavior. New York, NY: Appleton-Century-Crofts.

Hunt, L. T., Behrens, T. E. J., Hosokawa, T., Wallis, J. D., Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *eLife* 4, e11945.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb Cortex* 17, 2443–2452.

Jackson, S. A. W., Horst, N. K., Pears, A., Robbins, T. W., Roberts, A. C. (2016). Role of the perigenual anterior cingulate and orbitofrontal cortex in contingency learning in the marmoset. *Cereb Cortex* 26, 3273–3284.

Jocham, G., Brodersen, K. H., Constantinescu, A. O., Kahn, M. C., Ianni, A. M., Walton, M. E., *et al.* (2016). Reward-guided learning with and without causal attribution. *Neuron* 90, 177–190.

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., Graybiel, A. M. (1999). Building neural representations of habits. *Science* 286, 1745–1749.

Johnson, C. M., Peckler, H., Tai, L. H., Wilbrecht, L. (2016). Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nat Commun* 7, 10785.

Kaiser, R. H., Treadway, M. T., Wooten, D. W., Kumar, P., Goer, F., Murray, L., *et al.* (2017). Frontostriatal and dopamine markers of individual differences in reinforcement learning: a multi-modal investigation. *Cereb Cortex* 28(12), 4281–4290.

Kawai, T., Yamada, H., Sato, N., Takada, M., Matsumoto, M. (2015). Roles of the lateral habenula and anterior cingulate cortex in negative outcome monitoring and behavioral adjustment in nonhuman primates. *Neuron* 88, 792–804.

Kennerley, S. W., Behrens, T. E. J., Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience* 14, 1581–1589.

Kennerley, S. W., Dahmubed, A. F., Lara, A. H., Wallis, J. D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. *J Cogn Neurosci* 21, 1162–1178.

Khamassi, M., Enel, P., Dominey, P. F., Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. *Prog Brain Res* 202, 441–464.

Khamassi, M., Lallée, S., Enel, P., Procyk, E., Dominey, P. F. (2011). Robot cognitive control with a neurophysiologically inspired reinforcement learning model. *Front Neurorobot* 5, 1.

Khamassi, M., Quilodran, R., Enel, P., Dominey, P. F., Procyk, E. (2015). Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. *Cereb Cortex* 25, 3197–3218.

Kim, H., Lee, D., Jung, M. W. (2013). Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *J Neurosci* 33, 52–63.

Kim, H., Sul, J. H., Huh, N., Lee, D., Jung, M. W. (2009). Role of striatum in updating values of chosen actions. *J Neurosci* 29, 14701–14712.

Kötter, R., Wickens, J. (1995). Interactions of glutamate and dopamine in a computational model of the striatum. *J Comput Neurosci* 2, 195–214.

Lebedev, M. A., Messinger, A., Kralik, J. D., Wise, S. P. (2004). Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biol* 2, e365.

Mackintosh, N. J. (1975). Blocking of conditioned suppression: role of the first compound trial. *J Exp Psychol* 1, 335–345.

Mansouri, F. A., Matsumoto, K., Tanaka, K. (2006). Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *J Neurosci* 26, 2745–2756.

Markowitz, D. A., Curtis, C. E., Pesaran, B. (2015). Multiple component networks support working memory in prefrontal cortex. *Proc Natl Acad Sci U.S.A.* 112, 11084–11089.

Matsuda, Y., Marzo, A., Otani, S. (2006). The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. *J Neurosci* 26, 4803–4810.

McDannald, M. A., Esber, G. R., Wegener, M. A., Wied, H. M., Liu, T.-L., Stalnaker, T. A., *et al.* (2014). Orbitofrontal neurons acquire responses to "valueless" Pavlovian cues during unblocking. *eLife* 3, e02653.

Meyers, E. M., Qi, X. L., Constantinidis, C. (2012). Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proc Natl Acad Sci U.S.A.* 109, 4651–4656.

Mill, R. D., Bagic, A., Bostan, A., Schneider, W., Cole, M. W. (2017). Empirical validation of directed functional connectivity. *Neuroimage* 146, 275–287.

Monosov, I. E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nat Commun* 8, 134.

Montague, P. R., Dayan, P., Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16, 1936–1947.

Morrison, S. E., Saez, A., Lau, B., Salzman, C. D. (2011). Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* 71, 1127–1140.

Mulder, A. B., Nordquist, R. E., Orgüt, O., Pennartz, C. M. A. (2003). Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. *Behav Brain Res* 146, 77–88.

Niv, Y. (2009). Reinforcement learning in the brain. *J Math Psychol* 53, 139–154.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., et al. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *J Neurosci* 35, 8145–8157.

Noonan, M. P., Chau, B. K. H., Rushworth, M. F. S., Fellows, L. K. (2017). Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. *J Neurosci* 37, 7023–7035.

Noonan, M. P., Walton, M. E., Behrens, T. E., Sallet, J., Buckley, M. J., Rushworth, M. F. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc Natl Acad Sci U.S.A.* 107, 20547–20252.

Oemisch, M., Westendorff, S., Everling, S., Womelsdorf, T. (2015). Interareal spike-train correlations of anterior cingulate and dorsal prefrontal cortex during attention shifts. *J Neurosci* 35, 13076–13089.

Padoa-Schioppa, C., Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226.

Padoa-Schioppa, C., Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat Neurosci* 11, 95–102.

Pan, W. X., Schmidt, R., Wickens, J. R., Hyland, B. I. (2005). Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J Neurosci* 25, 6235–6242.

Pasupathy, A., Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433, 873–876.

Pearson, J. M., Hayden, B. Y., Raghavachari, S., Platt, M. L. (2009). Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. *Curr Biol* 19, 1532–1537.

Procyk, E., Tanaka, Y. L., Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nat Neurosci* 3, 502–508. 10.1038/74880

Quilodran, R., Rothe, M., Procyk, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* 57, 314–325.

Roelfsema, P. R., van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural Comput* 17, 2176–2214.

Rothkopf, C. A., Ballard, D. H. (2010). Credit assignment in multiple goal embodied visuomotor behavior. *Front Psychol* 1, 173.

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., et al. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J Neurosci* 28, 13775–13785.

Saez, R. A., Saez, A., Paton, J. J., Lau, B., Salzman, C. D. (2017). Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward. *Neuron* 95, 70.e3–77.e3.

Scholl, J., Klein-Flügge, M. (2017). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. *Behav Brain Res* 355, 56-75.

Schultz, W. (1998a). Predictive reward signal of dopamine neurons. *J Neurophysiol* 80, 1–27.

Schultz, W. (1998b). The phasic reward signal of primate dopamine neurons. *Adv Pharmacol* 42, 686–690.

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Curr Opin Neurobiol* 14, 139–147.

Schultz, W., Dickinson A. (2000). Neuronal coding of prediction errors. *Ann Rev Neurosci* 23, 473–500.

Seo, H., Barraclough, D. J., Lee, D. (2007). Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cerebral Cortex* 17(Suppl. 1), i110–i117.

Seo, H., Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27, 8366–8377.

Seo, M., Lee, E., Averbeck, B. B. (2012). Action selection and action value in frontal-striatal circuits. *Neuron* 74, 947–960.

Seol, G. H., Ziburkus, J., Huang, S., Song, L., Kim, I. T., Takamiya, K., *et al.* (2007). Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. *Neuron* 55, 919–929.

Shidara, M., Richmond, B. J. (2002). Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* 296, 1709–1711.

Shuler, M. G., Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science* 311, 1606–1609.

Sternson, S. M., Roth, B. L. (2014). Chemogenetic tools to interrogate brain functions. *Ann Rev Neurosci* 37, 387–407.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375.

Sul, J. H., Kim, H., Huh, N., Lee, D., Jung, M. W. (2010). Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* 66, 449–460.

Sutton, R. S., Barto, A. G. (1998). Reinforcement Learning: An Introduction Vol. 1 Cambridge: MIT press

Swanson, A. M., Allen, A. G., Shapiro, L. P., Gourley, S. L. (2015). GABAAα1-mediated plasticity in the orbitofrontal cortex regulates context-dependent action selection. *Neuropsychopharmacology* 40, 1027–1036.

Tremblay, L., Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature* 398, 704–708.

Tsujimoto, S., Genovesio, A., Wise, S. P. (2009). Monkey orbitofrontal cortex encodes response choices near feedback time. *J Neurosci* 29, 2569–2574.

Tsujimoto, S., Genovesio, A., Wise S. P. (2011). Comparison of strategy signals in the dorsolateral and orbital prefrontal cortex. *J Neurosci* 31, 4583–4592.

Urbanczik, R., Senn, W. (2009). Reinforcement learning in populations of spiking neurons. *Nat Neurosci* 12, 250–252.

Voloh, B., Valiante, T. A., Everling, S., Womelsdorf, T. (2015). Theta-gamma coordination between anterior cingulate and prefrontal cortex indexes correct attention shifts. *Proc Natl Acad Sci U.S.A.* 112, 8457–8462.

Wallis, J. D., Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur J Neurosci* 18, 2069–2081.

Walsh, M. M., Anderson, J. R. (2014). Navigating complex decision spaces: problems and paradigms in sequential choice. *Psychol Bull* 140, 466–486.

Wilson, R. C., Niv, Y. (2011). Inferring relevance in a changing world. *Front Hum Neurosci* 5, 189.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81, 267–279.

Womelsdorf, T., Ardid, S., Everling, S., Valiante, T. A. (2014). Burst firing synchronizes prefrontal and anterior cingulate cortex during attentional control. *Curr Biol* 24, 2613–2621.

Wörgötter, F., Porr, B. (2005). Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. *Neural Comput* 17, 245–319.

Xie, J., Padoa-Schioppa, C. (2016). Neuronal remapping and circuit persistence in economic decisions. *Nat Neurosci* 19, 855–861.

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends Cogn Sci* 21, 794–815.

Zhang, J. C., Lau, P.-M., Bi, G.-Q. (2009). Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. *Proc Natl Acad Sci U.S.A.* 106, 13028–13033.

Zsuga, J., Biro, K., Tajti, G., Szilasi, M. E., Papp, C., Juhasz, B., *et al.* (2016). 'Proactive' use of cue-context congruence for building reinforcement learning's reward function. *BMC Neurosci* 17, 70.

Chapter 5: Solving the credit assignment problem in adolescence

Abstract

Throughout adolescence, humans continue learning-to-learn. Previous studies have demonstrated that simple stimulus-action-reward learning skills are fully established by the end of childhood and that adolescents do not differ from adults in their baseline learning rates or in their ability to respond to abrupt changes in reward contingencies. However, adults appear to be better at learning from counterfactual feedback about non-chosen options and at calibrating their learning to reward statistics on the task than adolescents. The goal of the present work was to contribute to the rapidly growing area of investigation into developmental differences in reward-guided learning. Specifically, using a freely publicly available learning dataset (Palminteri, 2016), I compared adult and adolescent humans in their approaches to solving the *credit assignment problem* (i.e., the problem of discovering which choices are responsible for rewards). This study provides preliminary evidence for enhanced contingent credit assignment in adolescence: the general pattern of results suggests an enhanced sensitivity to contingent feedback and decreased sensitivity to non-contingent feedback in adolescents compared to adults. While adults integrate information about their previous decisions and past outcomes to guide their subsequent choices, adolescents are more narrowly focused on the most recent choice-reward history. This narrower behavioral focus on recency likely serves a functional purpose as it might better suit the needs of an adolescent in her rapidly changing physical and social environment.

Introduction

From early infancy, the young can learn to associate the stimuli in their environment and their own actions with reward, encouraging them to repeat behaviors that lead to positive reinforcement (e.g. a smile from a happy parent; Rovee and Rovee, 1969; Nussenbaum and Hartley, 2019). Across the stages of neural and behavioral development, reinforcement learning becomes more complex and adaptable: the brain learns-to-learn from secondary, abstract and probabilistic rewards, from counterfactual and hypothetical information, and in dynamic environments. By the end of childhood, simple stimulus-action-reward learning skills are fully established and adolescents do not differ from adults in their baseline learning rates or in their ability to respond to abrupt changes in reward contingencies (**Chapter 2**; Javadi et al., 2014). However, adults appear to be better at calibrating their learning to reward statistics on the task than adolescents (Nussenbaum and Hartley, 2019; Decker et al., 2015; Master et al., 2019). Compared to adolescents, adults are also more efficient at learning from counterfactual feedback about the non-chosen options and under conditions of negative framing (i.e., loss-avoidance framing; Palminteri et al., 2016).

To learn efficiently, an animal needs to discover which stimuli and actions are responsible for rewards (i.e., solve the *credit assignment problem*; Mackintosh, 1975). Several rules can be used to assign credit for rewards to previous choices: the causal relationship rule (i.e., contingent learning), the temporal proximity rule (i.e., when the reward is received shortly after a response), or the statistical rule (i.e., when an action has been executed frequently before beneficial outcomes; Jocham et al., 2016). Rewards have more influence on the behavior of adolescents than adults (Cauffman et al., 2010; Desor and Beauchamp, 1987; Friemel et al., 2010; Wilmouth and Spear, 2009; Doremus-Fitzwater and Spear, 2016). Furthermore, when presented with a choice between

a small immediate reward and large delayed reward, adolescents are more likely than adults to prefer the immediate option (Huang et al., 2017). The latter observation is usually interpreted as a deficit in impulse or cognitive control over behavior but could also indicate a difficulty in assigning credit for a temporally distal reward. It is presently unknown whether adolescents and adults differ in their approaches to solving the credit assignment problem.

The goal of the present work is to contribute to the rapidly growing area of investigation into developmental differences in reward-guided learning. Specifically, using a freely publicly available learning dataset (Palminteri, 2016), I compare adults and adolescents in their approaches to solving the credit assignment problem and their use of the Win-Stay and Lose-Shift strategies.

Methods

Dataset

The data analyzed in the present work are freely publicly available (Palminteri, 2016) and the methodological approach was described in detail in the associated publication (Palminteri et al., 2016).

Briefly, the analyses reported here are based on data from 20 adults (18-32 years old) and 18 adolescents (12-16 years old) tested on a probabilistic instrumental learning task. On each trial, the participants were offered a choice between two visual cues (characters from the Agathodaimon alphabet), each associated with a different outcome probability. There were eight cues in total, forming four stable cue pairs. Each of these pairs of cues defined a unique learning context differing in framing valence (reward framing vs loss framing) and information availability (complete vs partial).

Under reward framing conditions, the best of the two options led to a gain of 1 point with 0.75 probability and the worst option led to a gain of 1 point with 0.25 probability; on other trials, no points were earned. Under loss framing conditions, the worst option led to the loss of a point with 0.75 probability and the best option led to the loss of a point with 0.25 probability; no points were lost on other trials. Under partial information conditions a participant could only observe the outcome of the choice she made, whereas under complete information conditions she was also shown the outcome associated with the non-selected cue.

The participants were given 2 s to view the cues and make their selection by pressing the corresponding button, after which a red arrow briefly (0.5 s) pointed at the chosen option and the cues disappeared from the screen. Importantly, the outcome of the trial was only revealed after the cues were removed from the screen, imposing a temporal credit assignment requirement. Each

cue pair was presented 20 times in a pseudo-randomized order and spatial configuration for a total of 80 trials for each participant. To boost motivation, participants' compensation depended on the amount of points accumulated during the task. The primary study received ethics approval from the UCL Research Ethics Committee, and participants (adults), or their legal guardians (adolescents), gave written informed consent.

Win-Stay and Lose-Shift

Win-Stay, Lose-Shift, Counterfactual Win-Shift, and Counterfactual Loss-Stay probabilities were calculated based on trial-by-trial data. Each trial was classified as a win if the participant received the best possible outcome given the trial's valence frame (i.e., a point won for positively framed cue pairs and no point lost for negatively framed cue pairs). The decisions were classified as stays if the participant repeated the same choice during the next trial when the same cue pair was presented and as shifts if she switched to the other alternative. First, I analyzed the Win-Stay and Lose-Shift strategies for the outcomes of decisions made by participants (i.e., wins and losses associated with the chosen option) across all four unique learning contexts differing in framing valence and information availability. In these analyses, I calculated p(Win-Stay) and p(Lose-Shift) by dividing the number of Win-Stay and Lose-Shift trials by the total number of win and lose trials, respectively. Next, I asked whether adolescents and adults responded differently to counterfactual wins and losses, revealed for the non-chosen option under complete information conditions. For these analyses, I calculated p(Counterfactual Win-Shift) and p(Counterfactual Loss-Stay). The first of these probabilities, p(Counterfactual Win-Shift), reflects the frequency with which the participant switches to the previously unchosen option after being shown that it

resulted in a win; the second, p(Counterfactual Loss-Stay), reflects the frequency, with which she stays with her current choice after seeing that the unchosen option resulted in a loss.

Statistics

The data were analyzed with mixed-effects generalized linear models (GLM) in MATLAB (fitglme function; Statistics and Machine Learning Toolbox; MathWorks, Natick, Massachusetts; Versions R2017a and R2020a). Developmental group (adolescent vs adult) was modeled as a fixed factor. The analysis of percent correct also included trial block as a fixed factor and participant ID as a random factor. Importantly, I present data in overlapping sliding windows of trials in **Figure 1**, but the analyses only included two non-overlapping blocks (trials 1-10 and 11-20) because these statistical models are poorly suited to handling autocorrelated data. I judged the significance of each predictor based on the t-test of the associated beta coefficient. Statistical significance was noted when p-values were less than 0.05.

Credit Assignment

Similar to previous investigations (Jocham et al., 2016; Noonan et al., 2017; Barraclough et al., 2004; Lau and Glimcher, 2007; Walton et al., 2011), I examined how previous decisions and their outcomes influence choices on future trials. Given the paucity of data available for each participant, the typically used models that incorporate 4-trial-long reward histories and estimate unique parameters for each subject failed to converge. Therefore, the present analyses considered shorter-term choice and reward histories, up to 2 trials into the past, and combined the trial-by-trial data across participants within each developmental group. I fit GLM to choices of participants in each unique learning context to estimate the combined influence of previous choices and

associated rewards on current behavior. The outcome variable in each case was the participant's

choice on trial t and the predictor variables were combinations of recently made decisions and their

outcomes (on t-1 and t-2) in the same learning context. I present the beta coefficients as matrices

in **Figure 5**. In each case, the diagonal of the matrix represents contingent learning. As has been

observed before, outcomes of behavior can frequently reinforce temporarily adjacent, yet

unrelated, choices. Alternatively, the recent history of rewards can be jointly attributed to the most

recent choice (Thorndike 1933; Noonan et al., 2010; Walton et al., 2011; Jocham et al., 2016).

These influences are represented in the off-diagonal elements of each matrix. Like in the general

statistical approach, I judged the significance of each predictor based on the t-test of the associated

beta coefficient. Any developmental group differences were revealed by significant beta

coefficients for the choice x reward x group interactions. Statistical significance was noted when

p-values were less than 0.05.

Results

Developmental differences in learning

Under partial information conditions of the reward framing context, adolescents demonstrated better overall performance compared to adults [**Figure 5-1**; main effect of developmental group: β=0.19912, t(72)=2.0896, p=0.040183]. This performance enhancement in adolescents appeared to be most prevalent during the first half of learning, with adults demonstrating a greater improvement across learning blocks. The analyses, however, were not sufficiently powered to detect the block x group interaction for this effect [β=-0.11121, t(72)=-1.392, p=0.1682]. Under complete information conditions of the reward framing context, there was a marginally significant block x group interaction [β=-0.12017, t(72)=-1.7989, p=0.076229], with adolescents showing a lower rate of learning across trial blocks compared to adults.

There were no significant developmental group differences in learning under partial information conditions of the loss framing context (**Figure 5-1**): the analyses revealed neither a significant main effect of group [β=0.057857, t(72)=0.4717, p=0.63856], nor a significant block x group interaction [β=-0.084172, t(72)=-1.1234, p=0.26498]. Under complete information conditions of the loss framing context, there was a marginally significant main effect of developmental group with adolescents demonstrating worse performance compared to adults [β=-0.23794, t(72)=-1.6865, p=0.09602]. Overall, adolescents performed slightly better than adults in the partial information/reward framing context but learned at a lower rate under complete information conditions in the reward framing context and performed worse in the complete information/loss framing context. These results further validate the findings of the primary study using a different statistical approach (Palminteri et al., 2016).
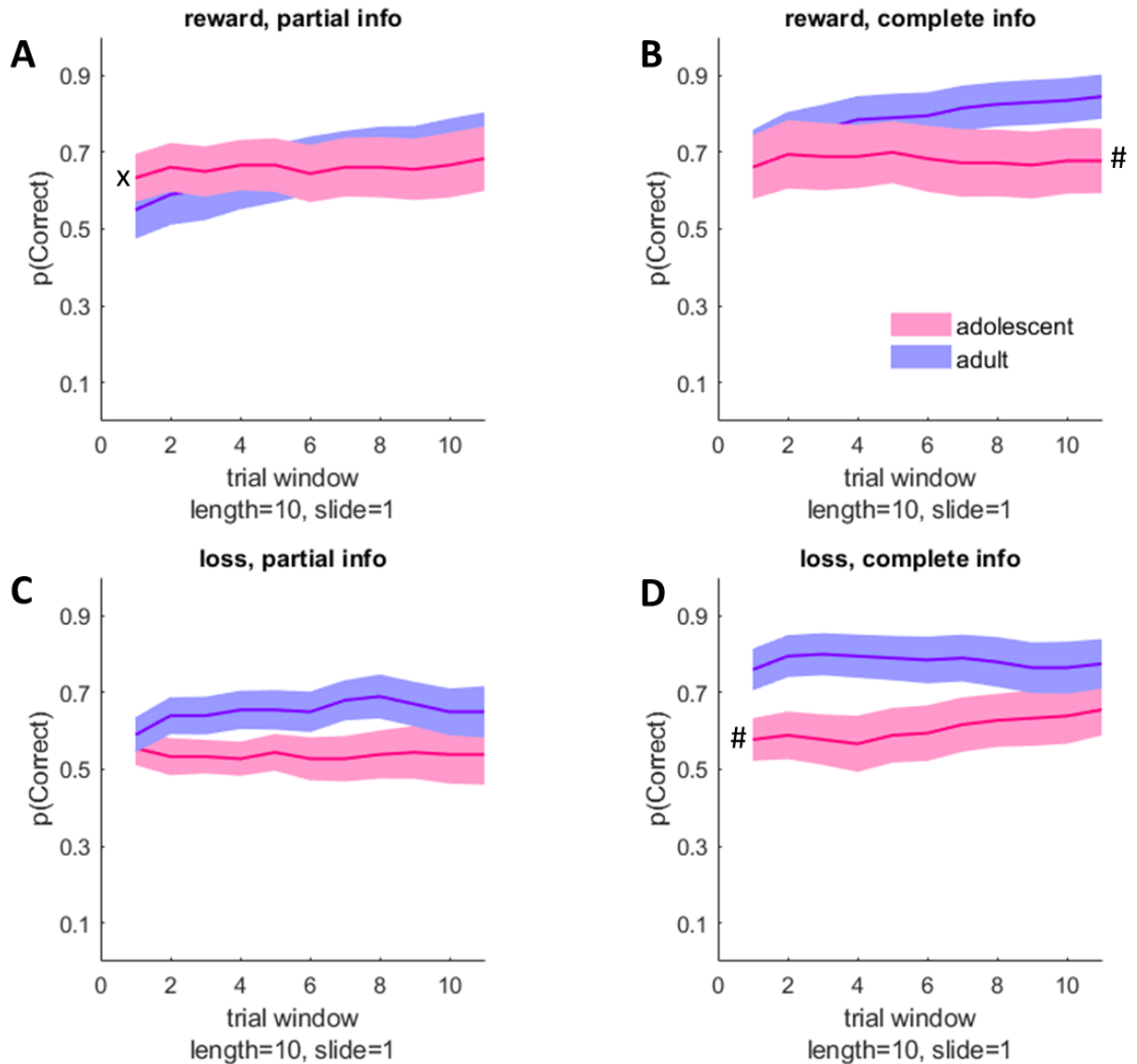
Figure 5-1. Developmental differences in learning across contexts varying in information availability and framing valence. A. Under partial information conditions of the reward framing context, adolescents demonstrated a slightly better overall performance compared to adults. **B.** Under complete information conditions of the reward framing context, there was a marginally significant block x group interaction (p=0.076), with adolescents showing a lower rate of learning across trial blocks compared to adults. **C.** There were no significant developmental group differences in learning under partial information conditions of the loss framing context. **D.** Under complete information conditions of the loss framing context, there was a marginally significant main effect of developmental group with adolescents demonstrating worse performance overall compared to adults (p=0.096). Solid lines represent group averages of probability correct (i.e., best option chosen given the context) in sliding windows of 10 trials with the slide of 1 trial. Shaded areas represent SEM. x<0.05, 0.05<#<0.1. x and # on the left of the graph indicate the significance of the main effect and # on the right of the graph indicates the significance of the interaction.

Win-Stay Lose-Shift strategies

First, I analyzed Win-Stay and Lose-Shift strategies for the outcomes of decisions made by participants (i.e., wins and losses associated with the chosen option). There was a marginally significant main effect of developmental group under partial information conditions of the reward framing context [**Figure 5-2**; β=-0.13372, t(36)=-1.8463, p=0.073076] and a significant main effect of developmental group under complete information conditions of the loss framing context [β=-0.18724, t(36)=-3.6474, p=0.00083177] for the Win-Stay strategy. In both cases, adolescents were less likely to repeat the same choice after winning during their next encounter with the same learning context compared to their adult counterparts. While adolescents also appeared to use the Win-Stay strategy less under complete information conditions of the reward framing context [β=-0.10815, t(36)=-1.5807, p=0.1227] and under partial information conditions of the loss framing context [β=-0.068713, t(36)=-1.196, p=0.23951], these differences were not statistically significant. Unlike Win-Stay, the Lose-Shift strategy was not affected by developmental stage (**Figure 5-3**): partial information conditions of the reward framing context [β=-0.013466, t(36)=-0.18919, p=0.85101], complete information conditions of the reward framing context [β=-0.029933, t(36)=-0.48828, p=0.62831], partial information conditions of the loss framing context [β=-0.015658, t(36)=-0.28983, p=0.77361], or complete information conditions of the loss framing context [β=0.077798, t(36)=1.6098, p=0.11617].
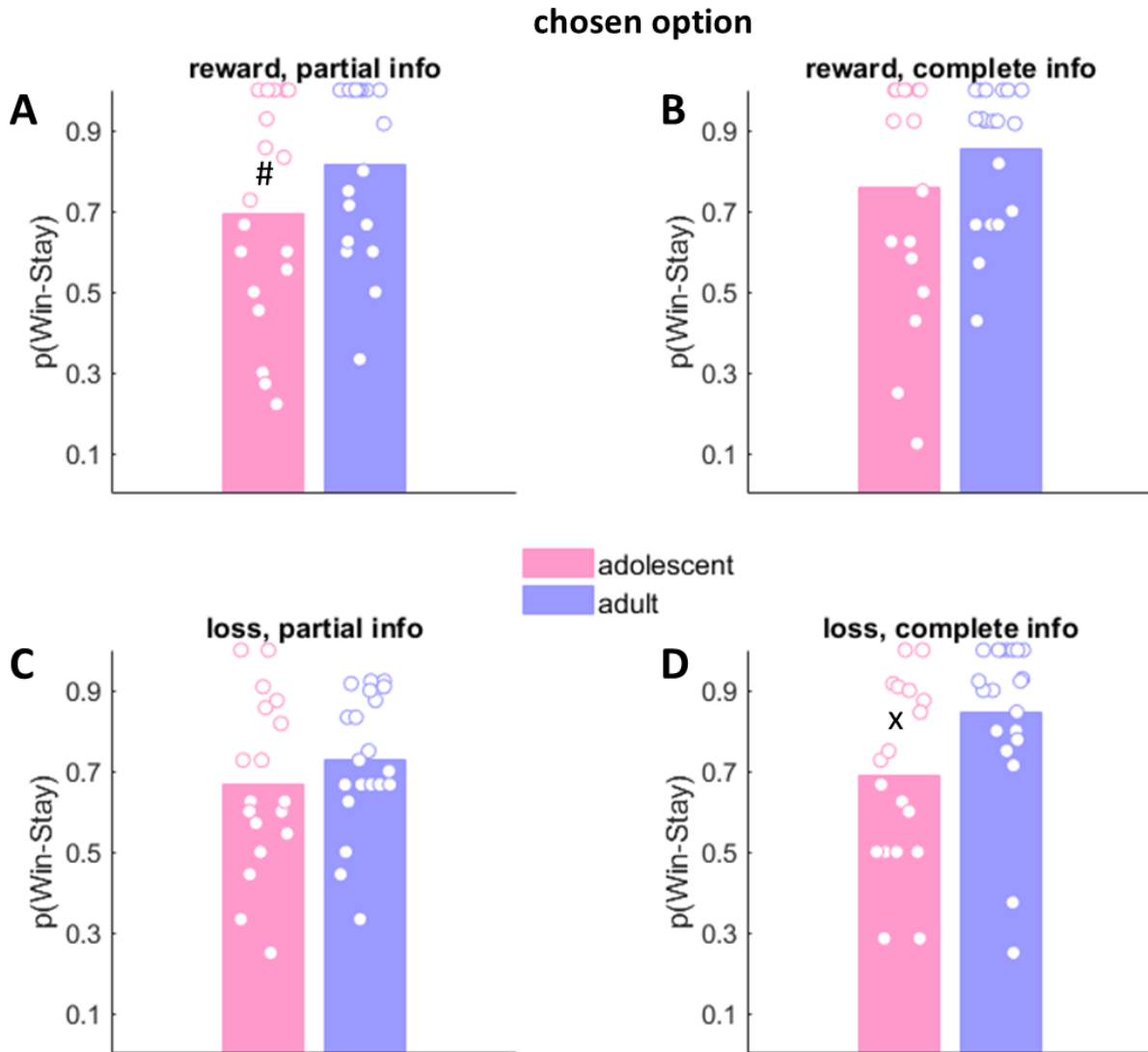
Figure 5-2. Developmental differences in Win-Stay strategy use. A, D. Adolescents were marginally less likely to use the Win-Stay strategy under partial information condition of the reward framing context (p=0.073) and significantly less likely to use this strategy under complete information conditions of the loss framing context compared to adults. **B, C.** There were no significant differences in the use of Win-Stay strategy between adolescents and adults under complete information conditions of the reward framing context or under partial information conditions of the loss framing context. Bars represent group averages and dots represent individual participant data. x<0.05, 0.05<#<0.1.
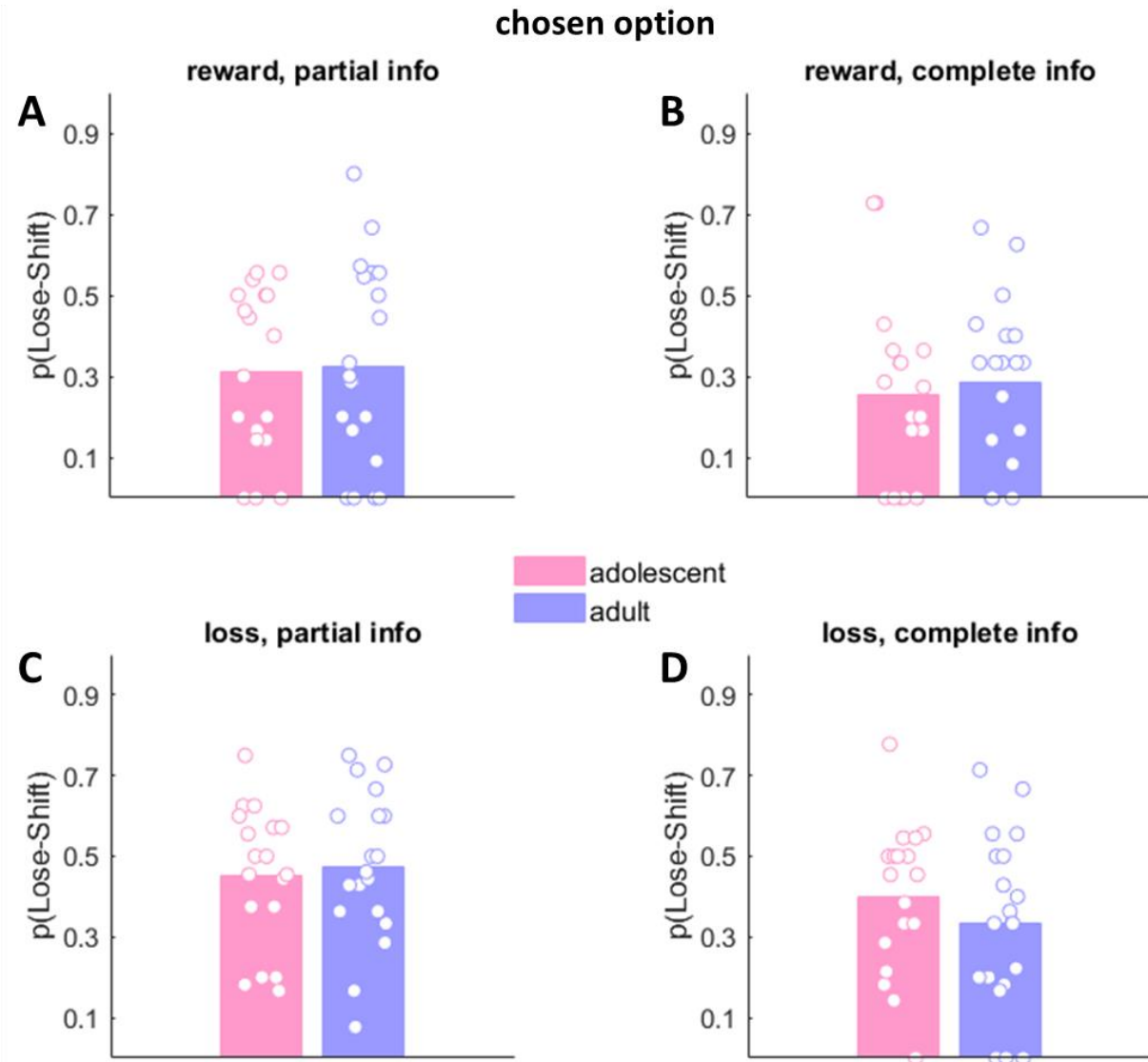
Figure 5-3. Developmental differences in Lose-Shift strategy use. A-D. The Lose-Shift strategy was not affected by developmental stage: there were no developmental group differences in any of the learning contexts. Bars represent group averages and dots represent individual participant data.
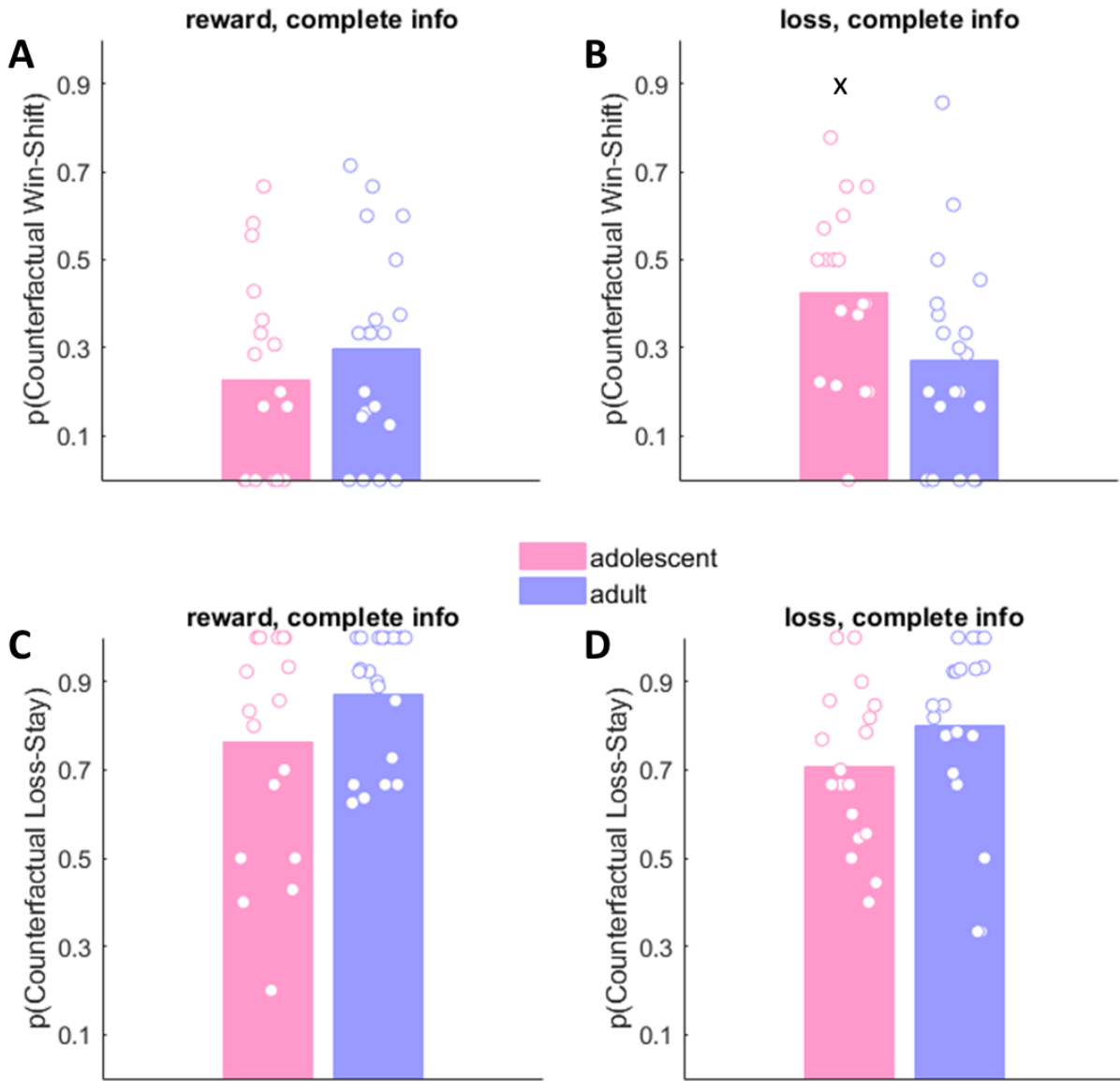
Figure 5-4. Developmental differences in the use of counterfactual Win-Shift and counterfactual Loss-Stay strategies. A. There were no developmental differences in the tendency to switch away from the current choice after observing a counterfactual win in the reward framing context**. B.** In the loss framing context, adolescents were significantly more likely to use the counterfactual Win-Shift strategy. **C, D.** There were no developmental group differences in the tendency to stay with the currently chosen option after observing a counterfactual loss either in the reward or loss framing context. Bars represent group averages and dots represent individual participant data. x<0.05.

Second, I analyzed the responses to counterfactual wins and losses, revealed for the non-chosen option under complete information conditions. While adolescents appeared to be more likely to stay with their chosen option after observing counterfactual losses, these differences were not statistically significant under complete information conditions of the reward [β=0.044923, t(36)=0.64145, p=0.52529] or loss [β=0.045636, t(36)=0.82733, p=0.4135] framing contexts (**Figure 5-4**). Under complete information conditions of the loss framing context, adolescents were significantly more likely compared to their adult counterparts to switch away from their previously chosen option after observing a counterfactual win [**Figure 5-4**; β=0.30309, t(36)=5.479, p=3.455*10^{-06}]. For the reward framing context, these analyses detected no significant differences in the probability of switching after observing counterfactual wins [β=0.086354, t(36)=1.5714, p=0.12484].

Developmental differences in credit assignment

I fit GLM to choices of participants in each unique learning context to estimate the combined influence of previous choices and associated rewards (up to 2 trials into the past, t-1 and t-2) on current behavior (the choice at trial t). Generally, these analyses revealed the typical pattern of sensitivity to previous choice-outcome histories in both age groups: the combined influence of choice at t-1 and reward at t-1 had the strongest influence on current behavior (**Figure 5-5**). The notable exception are the estimates for the partial information/reward framing condition. Here, adults appear to be learning more from the non-contingent combinations of choices and rewards and adolescents are more sensitive to longer-term than to recent choice-outcome histories. While I still present the results for this condition, I caution against their overinterpretation, because the

aforementioned peculiarities could be due to the small number of observations on which the beta coefficients were estimated, potentially leading to unreliability in their estimation.

For the partial information/reward framing condition, GLM analyses revealed increased sensitivity to contingent feedback in adolescents compared to adults [$choice_{t-1}$ x $reward_{t-1}$: $\beta=0.35649$, $t(675)=2.3028$, $p=0.021594$; $choice_{t-2}$ x $reward_{t-2}$: $\beta=0.86535$, $t(675)=4.2803$, $p=2.1364*10^{-05}$] and decreased sensitivity to non-contingent feedback in adolescents compared to adults [$choice_{t-2}$ x $reward_{t-1}$: $\beta=-0.67581$, $t(675)=-4.3437$, $p=1.6167*10^{-05}$; $choice_{t-1}$ x $reward_{t-2}$: $\beta=-0.58857$, $t(675)=-2.8232$, $p=0.0048946$]. For the complete information/reward framing condition, GLM analyses also revealed increased sensitivity to contingent feedback in adolescents compared to adults, but only for choice-reward history 2 trials into the past [$choice_{t-2}$ x $reward_{t-2}$: $\beta=0.34516$, $t(675)=2.0159$, $p=0.044204$]. For the partial information/loss framing condition, the analyses revealed reduced sensitivity to non-contingent choice-reward history, specifically the combined influence of choices at t-1 and rewards at t-1, in adolescents [$choice_{t-2}$ x $reward_{t-1}$: $\beta=-0.45327$, $t(675)=-2.4985$, $p=0.012708$]. Finally, GLM analyses revealed no developmental group differences in credit assignment for the complete information/loss framing condition. Overall, despite the differences across conditions the general pattern of the results suggests that contingent credit assignment is potentiated in adolescence: compared to adults, adolescents are more sensitive to contingent feedback and less sensitive to non-contingent feedback on previous choices when making decisions.
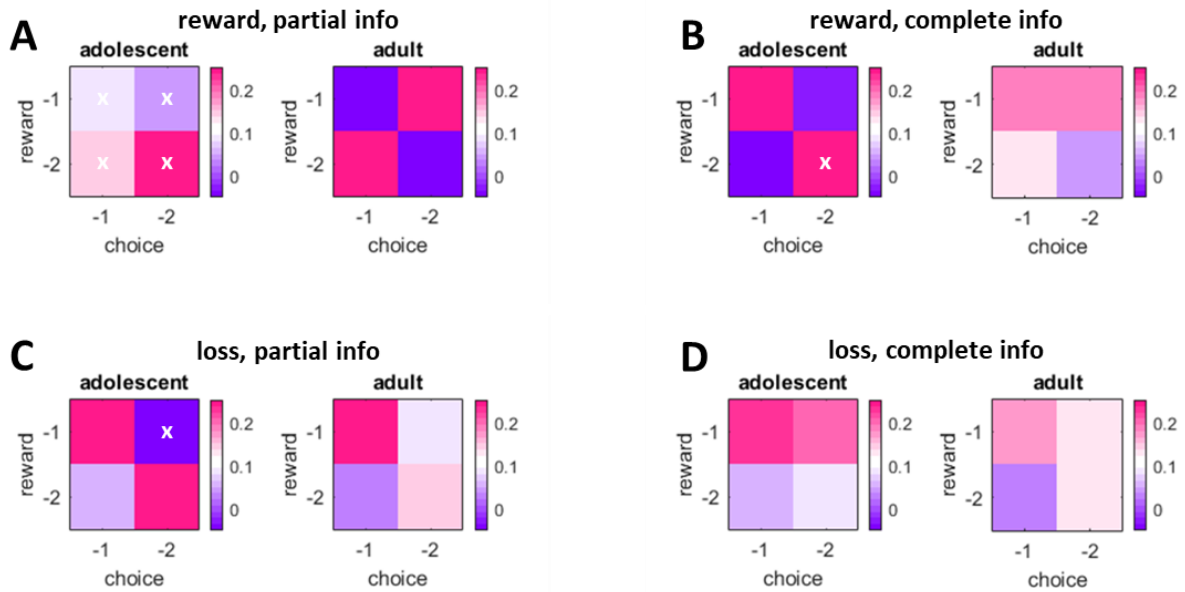
Figure 5-5. Developmental differences in credit assignment. A. For the partial information/reward framing condition, GLM analyses revealed increased sensitivity to contingent feedback in adolescents compared to adults [choice$_{t-1}$ x reward$_{t-1}$: p=0.02; choice$_{t-2}$ x reward$_{t-2}$: p=2.1364*10^{-05}] and decreased sensitivity to non-contingent feedback in adolescents compared to adults [choice$_{t-2}$ x reward$_{t-1}$: p=1.6167*10^{-05}; choice$_{t-1}$ x reward$_{t-2}$: p=0.005]. **B.** For the complete information/reward framing condition, GLM analyses also revealed increased sensitivity to contingent feedback in adolescents compared to adults, but only for choice-reward history 2 trials into the past [choice$_{t-2}$ x reward$_{t-2}$: p=0.04]. **C.** For the partial information/loss framing condition, the analyses revealed reduced sensitivity to non-contingent choice-reward history, specifically the combined influence of choices at t-2 and rewards at t-1, in adolescents [choice$_{t-2}$ x reward$_{t-1}$: p=0.01]. **D.** GLM analyses revealed no developmental group differences in credit assignment for the complete information/loss framing condition. The color of each matrix element represents the magnitude of the β coefficient that reflects the combined influence of previous choices and associated rewards on current behavior. The diagonal elements of each matrix correspond to contingent credit assignment. x<0.05.

118

Discussion

This study provides preliminary evidence for enhanced contingent credit assignment in adolescence. While findings varied across learning contexts differing in information availability and framing valence, the general pattern of results suggests an enhanced sensitivity to contingent feedback and decreased sensitivity to non-contingent feedback in adolescents compared to adults. The results also suggest that adults integrate information about their previous decisions and past outcomes when making subsequent choices, while adolescents are more narrowly focused on the most recent choice-reward combination. Across all unique learning contexts, the magnitude of the beta coefficient for $choice_{t-1}$ x $reward_{t-1}$ was greater for adolescents compared to adults (note that this difference reached statistical significance only in the partial information/reward framing context); this occurred despite the lower p(Win-Stay) in adolescents, which at first may appear counterintuitive. Note, however, that all combinations of recently made decisions and their outcomes were included as predictors in the GLM for credit assignment analyses, and as such each beta coefficient represents the influence of a given choice-reward pair after accounting for the effects of other predictors. In adolescents, the relative magnitude of the $choice_{t-1}$ x $reward_{t-1}$ beta coefficient is larger than those of other predictors, demonstrating a narrow behavioral focus. In contrast, the beta coefficients in the adult age group were relatively uniform, indicating that adults are guided by longer-term and non-contingent choice-reward history to a greater degree than adolescents when making decisions.

Successful solution to the credit assignment problem and contingent learning have been linked to the function of the orbitofrontal cortex in humans (Jocham et al., 2016; Noonan et al., 2017) and monkeys (Noonan et al., 2010). Animals with lesions to the orbitofrontal cortex are unable to associate a reward with the choice on which it was contingent. Recent accounts further

suggest that contingent credit assignment relies on functional interactions between different subregions within the prefrontal cortex, including the orbitofrontal, anterior cingulate, and dorsolateral prefrontal cortices (Khamassi et al., 2011, 2013, 2015; Stolyarova, 2018). These cortical regions continue developing throughout adolescence and into adulthood (Geier et al., 2010; Sturman and Moghaddam, 2011; Ernst, 2014; Simon and Moghaddam, 2015; Sowell et al., 1999; Galván et al., 2006). The orbitofrontal cortex is more excited by reward receipt in adolescent rats compared to adults, which may contribute to the enhanced sensitivity to contingent feedback in this age group (Sturman and Moghaddam, 2011).

The findings on adolescent sensitivity to positive and negative feedback during learning have varied greatly from study to study, with investigators reporting lesser, greater, or equal feedback sensitivity in adolescents compared to adults (DePasque and Galván, 2019; van den Bos et al., 2012; van den Bos et al., 2009; DePasque and Galván, 2017; Humphreys et al., 2016; van Duijvenvoorde et al., 2008; Zhuang et al., 2017; Hauser et al., 2015; van der Schaff et al, 2011). Even within the current study, the effects of developmental stage on Win-Stay, Lose-Shift, Win-Shift and Loss-Stay strategies appear difficult to reconcile. The results suggest that adolescents are as sensitive to losses as adults: there were no developmental group differences in the use of the Lose-Shift strategy or counterfactual Loss-Stay strategy across learning contexts. The adolescents were, however, less likely than their adult counterparts to use the Win-Stay strategy in the partial information/reward framing context and complete information/loss framing context, suggesting a *reduced* sensitivity to positive feedback in this developmental group. At the same time, adolescents were more likely than adults to shift towards the previously unchosen option after observing a counterfactual win in the complete information/reward framing context, suggesting an *increased* sensitivity to positive feedback in this developmental group. These divergent findings across

120

studies and within the same group of participants tested on the same task are especially puzzling, given the well-documented phenomenon of enhanced reward-sensitivity in adolescence (van Duijvenvoorde et al., 2016; Doremus-Fitzwater et al., 2010; Steinberg et al., 2009; Urosevic et al., 2012; Andersen et al., 2002; Doremus-Fitzwater and Spear, 2016).

If the neural responses to positive feedback (in the form of reward prediction error) peak in adolescence (Cohen et al., 2010), why do they not reliably drive an increase in Win-Stay? And why are disagreements on adolescent sensitivity to negative feedback so common? I suggest that the discordant developmental effects on sensitivity to positive and negative feedback reported in the literature arise from the inherent increase in variability within the adolescent group that results from a combination of task and individual factors. As reinforcement learning is still developing during adolescence, task parameters, for example the reinforcement schedule, will influence adolescent performance (Nussenbaum and Hartley, 2019; Decker et al., 2015; Master et al., 2019).

On the individual level, the two most relevant factors for feedback learning are the current stage of neural development and the previous experience and level of practice, - factors that are greatly interrelated. On simpler task variants with deterministic reward contingencies, subcortical structures tend to support the use of Win-Stay and Lose-Shift strategies (e.g. the striatum; Thapa and Gruber, 2018; Aquili et al., 2014). Conversely, on more demanding tasks with probabilistic or rapidly changing reward contingencies, the frontocortical regions tend to play a greater role (Chau et al., 2015; Paulus et al., 2002; Constantinople et al., 2019; Stolyarova and Izquierdo, 2017; Stolyarova, Rakhshan et al., 2019; Riceberg and Shapiro, 2012). As mentioned above, the development of the frontal cortex and its dopaminergic innervation follow a delayed time course in adolescence (Geier et al., 2010; Sturman and Moghaddam, 2011; Ernst, 2014; Simon and Moghaddam, 2015; Sowell et al., 1999; Galván et al., 2006; Benes et al., 2000; Andersen et al.,

2000; Andersen et al., 2002; Naneix et al., 2012; Weickert et al., 2007), suggesting that the reduced use of Win-Stay or Lose-Shift strategies might be expected on some complex tasks in this age group. As for the influence of individual experience, it has been previously shown that a learning deficit in adolescence can be alleviated by practice with the task earlier in life (Garske et al., 2013). These findings suggest that an adolescent's prior experience with learning tasks may interact with her neural development to guide future reward responses and decisions.

Throughout adolescence, we continue learning-to-learn. The present results contribute to the rapidly growing area of investigation into developmental differences in reinforcement learning by showing an enhancement in contingent credit assignment in the adolescent group. While adults integrate information about their previous decisions and past outcomes to guide their subsequent choices, adolescents are more narrowly focused on the most recent choice-reward history. This narrower behavioral focus on recency likely serves a functional purpose as it might better suit the needs of an adolescent in her rapidly changing physical and social environment.

References

Andersen, S. L., Thompson, A. P., Krenzel, E., Teicher, M. H. (2002). Pubertal changes in gonadal hormones do not underlie adolescent dopamine receptor overproduction. *Psychoneuroendocrinology* 27, 683–691.

Andersen, S. L., Thompson, A. T., Rutstein, M., Hostetter, J. C., Teicher, M. H. (2000). Dopamine receptor pruning in prefrontal cortex during the periadolescent period in rats. *Synapse* 37, 167–169.

Aquili, L., Liu, A. W., Shindou, M., Shindou, T., Wickens, J. R. (2014). Behavioral flexibility is increased by optogenetic inhibition of neurons in the nucleus accumbens shell during specific time segments. *Learn Mem* 21(4), 223-31.

Barraclough, D. J., Conroy, M. L., Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7(4), 404-10.

Benes, F. M., Taylor, J. B., Cunningham, M. C. (2000). Convergence and plasticity of monoaminergic systems in the medial prefrontal cortex during the postnatal period: implications for the development of psychopathology. *Cereb Cortex* 10, 1014–1027.

Cauffman, E., Shulman, E. P., Steinberg, L., Claus, E., Banich, M. T., Graham, S., et al. (2010). Age differences in affective decision making as indexed by performance on the Iowa Gambling Task. *Dev Psychol* 46, 193–207.

Chau, B. K., Sallet, J., Papageorgiou, G. K., Noonan, M. P., Bell, A. H., Walton, M. E., et al. (2015). Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. *Neuron* 87(5), 1106-18.

Cohen, J. R., Asarnow, R. F., Sabb, F. W., Bilder, R. M., Bookheimer, S. Y., Knowlton, B. J., et al. (2010). A unique adolescent response to reward prediction errors. *Nature Neuroscience* 13(6), 669-671.

Constantinople, C. M., Piet, A. T., Bibawi, P., Akrami, A., Kopec, C., Brody, C. D. (2019). Lateral orbitofrontal cortex promotes trial-by-trial learning of risky, but not spatial, biases. *Elife* 8, e49744.

Decker, J. H., Lourenco, F. S., Doll, B. B., Hartley, C. A. (2015). Experiential reward learning outweighs instruction prior to adulthood. *Cogn Affect Behav Neurosci* 15(2), 310–320

DePasque, S., Galván, A. (2017). Frontostriatal development and probabilistic reinforcement learning during adolescence. *Neurobiol Learn Mem* 143, 1-7.

DePasque, S., Galván, A. (2019). Neurobiological responses in the adolescent striatum to being 'tested'. *Soc Cogn Affect Neurosci* 14(1), 3-12.

Desor, J. A., Beauchamp, G. K. (1987). Longitudinal changes in sweet preferences in humans. *Physiol Behav* 39, 639–641.

Doremus-Fitzwater, T. L., Spear, L. P. (2016). Reward-centricity and attenuated aversions: An adolescent phenotype emerging from studies in laboratory animals. *Neurosci Biobehav Rev* 70, 121–134.

Doremus-Fitzwater, T. L., Varlinskaya, E. I., Spear, L. P. (2010). Motivational systems in adolescence: possible implications for age differences in substance abuse and other risk-taking behaviors. *Brain Cogn* 72, 114–123.

Ernst, M. (2014). The triadic model perspective for the study of adolescent motivated behavior. *Brain Cognit* 89, 104–111.

Friemel, C. M., Spanagel, R., Schneider, M. (2010). Reward sensitivity for a palatable food reward peaks during pubertal developmental in rats. *Front Behav Neurosci* 4.

Galván, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., et al. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *J Neurosci* 26, 6885–6892.

Garske, A. K., Lawyer, C. R., Peterson, B. M., and Illig, K. R. (2013). Adolescent changes in dopamine D1 receptor expression in orbitofrontal cortex and piriform cortex accompany an associative learning deficit. *PLoS One* 8:e56191.

Geier, C. F., Terwilliger, R., Teslovich, T., Velanova, K., Luna, B. (2010). Immaturities in reward processing and its influence on inhibitory control in adolescence. *Cereb Cortex* 20, 1613–1629.

Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D., Brem, S. (2015). Cognitive flexibility in adolescence: neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *Neuroimage* 104, 347-54.

Huang, Y., Hu, P., Li, X. (2017). Undervaluing delayed rewards explains adolescents' impulsivity in inter-temporal choice: an ERP study. *Sci Rep* 7, 42631.

Humphreys, K. L., Telzer, E. H., Flannery, J., Goff, B., Gabard-Durnam, L., Gee, D. G., et al. (2016). Risky decision making from childhood through adulthood: Contributions of learning and sensitivity to negative feedback. *Emotion* 16(1), 101-9.

Javadi, A. H., Schmidt, D. H. K., Smolka, M. N. (2014). Adolescents adapt more slowly than adults to varying reward contingencies. *J Cogn Neurosci* 26(12), 2670–2681.

Jocham, G., Brodersen, K. H., Constantinescu, A. O., Kahn, M. C., Ianni, A. M., Walton, M. E., *et al.* (2016). Reward-guided learning with and without causal attribution. *Neuron* 90, 177–190.

Khamassi, M., Enel, P., Dominey, P. F., Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. *Prog Brain Res* 202, 441–464.

Khamassi, M., Lallée, S., Enel, P., Procyk, E., Dominey, P. F. (2011). Robot cognitive control with a neurophysiologically inspired reinforcement learning model. *Front Neurorobot* 5, 1.

Khamassi, M., Quilodran, R., Enel, P., Dominey, P. F., Procyk, E. (2015). Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. *Cereb Cortex* 25, 3197–3218.

Lau, B., Glimcher, P. W. (2007). Action and outcome encoding in the primate caudate nucleus. *J Neurosci* 27(52), 14502-14.

Mackintosh, N. J. (1975). Blocking of conditioned suppression: role of the first compound trial. *J Exp Psychol* 1, 335–345.

Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., Collins, A. G. E. (2020). Distentangling the systems contributing to changes in learning during adolescence. *Dev Cogn Neurosci* 41:100732.

Naneix, F., Marchand, A. R., Di Scala, G., Pape, J. R., Coutureau, E. (2012). Parallel maturation of goal-directed behavior and dopaminergic systems during adolescence. *J Neurosci* 32, 16223–16232

Noonan, M. P., Chau, B. K. H., Rushworth, M. F. S., Fellows, L. K. (2017). Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. *J Neurosci* 37, 7023–7035.

Noonan, M. P., Walton, M. E., Behrens, T. E., Sallet, J., Buckley, M. J., Rushworth, M. F. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc Natl Acad Sci U.S.A.* 107, 20547–20252.

Nussenbaum, K., Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Dev Cogn Neurosci*. 40: 100733.

Palminteri, S. (2016). Behavioural data and code the plot them. *figshare*. Dataset. https://doi.org/10.6084/m9.figshare.3398056.v2

Palminteri, S., Kilford, E. J., Coricelli, G., Blakemore, S. J. (2016). The Computational Development of Reinforcement Learning during Adolescence. *PLoS Comput Biol* 12(6), e1004953.

Paulus, M. P., Hozack, N., Frank, L., Brown, G. G. (2002). Error rate and outcome predictability affect neural activation in prefrontal cortex and anterior cingulate during decision-making. *Neuroimage* 15(4), 836-46.

Riceberg, J. S., Shapiro, M. L. (2012). Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *J Neurosci* 32(46), 16402-9.

Rovee, C. K., Rovee, D. T. (1969). Conjugate reinforcement of infant exploratory behavior. *J Exp Child Psychol* 8(1), 33–39.

Simon, N. W., Moghaddam, B. (2015). Neural processing of reward in adolescent rodents. *Dev Cogn Neurosci* 11, 145–154.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nat Neurosci* 2, 859–861.

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., Banich, M. (2009). Age differences in future orientation and delay discounting. *Child Dev* 80, 28–44.

Stolyarova, A. (2018). Solving the credit assignment problem with the prefrontal cortex. *Front Neurosci* 2, 182.

Stolyarova, A., Izquierdo, A. (2017). Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning and decision making under uncertainty. *Elife* 6. pii: e27483.

Stolyarova, A., Rakhshan, M., Hart, E. E., O'Dell, T. J., Peters, M. A. K., Lau, H., et al. (2019). Dissociable roles for anterior cingulate cortex and basolateral amygdala in decision confidence and learning under uncertainty. *Nature Communications* 10(1), 4704.

Sturman, D. A., Moghaddam, B. (2011). The neurobiology of adolescence: changes in brain architecture, functional dynamics, and behavioral tendencies. *Neurosci Biobehav Rev* 35, 1704–1712.

Sturman, D. A., Moghaddam, B. (2011). Reduced neuronal inhibition and coordination of adolescent prefrontal cortex during motivated behavior. *J Neurosci* 31, 1471–1478.

Thapa, R., Gruber, A. J. (2018). Lesions of ventrolateral striatum eliminate lose-shift but not win-stay behaviour in rats. *Neurobiol Learn Mem* 155, 446-451.

Thorndike, E. L. (1933). THE "SPREAD" OR "SCATTER" OF THE INFLUENCE FROM A REWARD, IN RELATION TO GESTALT DOCTRINES. *Science* 77(1998), 368.

Urosevic, S., Collins, P., Muetzel, R., Lim, K., Luciana, M. (2012). Longitudinal changes in behavioral approach system sensitivity and brain structures involved in reward processing during adolescence. *Dev Psychol* 48, 1488–1500.

van den Bos, W., Cohen, M. X., Kahnt, T., Crone, E. A. (2012). Striatum-Medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral Cortex* 22(6), 1247-1255

van den Bos, W., Güroğlu, B., van den Bulk, B. G., Rombouts, S. A. R. B., Crone, E. A. (2009). Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. *Frontiers in Human Neuroscience* 3, 52

van der Schaaf, M. E., Warmerdam, E., Crone, E. A., Cools, R. (2011). Distinct linear and non-linear trajectories of reward and punishment reversal learning during development: relevance for dopamine's role in adolescent decision making. *Dev Cogn Neurosci* 1(4), 578-90.

van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E., Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *J Neurosci* 28(38), 9495-94503.

van Duijvenvoorde, A. C., Peters, S., Braams, B. R., Crone, E. A. (2016). What motivates adolescents? Neural responses to rewards and their influence on adolescents' risk taking, learning, and cognitive control. *Neurosci Biobehav Rev* 70, 135-147.

Walton, M. E., Behrens, T. E., Noonan, M. P., Rushworth, M. F. (2011). Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world. *Ann N Y Acad Sci* 1239, 14-24.

Weickert, C. S., Webster, M. J., Gondipalli, P., Rothmond, D., Fatula, R. J., Herman, M. M., et al. (2007). Postnatal alterations in dopaminergic markers in the human prefrontal cortex. *Neuroscience* 144, 1109–1119.

Wilmouth, C. E., Spear, L. P. (2009). Hedonic sensitivity in adolescent and adult rats: taste reactivity and voluntary sucrose consumption. *Pharmacol Biochem Behav* 92, 566–573.

Zhuang, Y., Feng, W., Liao, Y. (2017). Want More? Learn Less: Motivation Affects Adolescents Learning from Negative Feedback. *Front Psychol* 8, 76.

Chapter 6: Conclusions

Summary

In **Chapter 2**, the data demonstrate that adolescent rats do not differ from adults in a simple form of stimulus-reward learning but show an enhanced motivation to invest effort to obtain larger rewards. Our findings that adolescent rats are more likely to climb barriers of a greater height to obtain larger rewards are congruent with a previous report of adolescent animals tolerating higher effort requirements on a progressive ratio lever pressing task (Friemel et al., 2010) and a study showing that adolescent humans are willing to expend greater physical effort in form of button presses during goal pursuit (Sullivan-Toole et al., 2019). While many previous studies have focused on the role of neurodevelopmental changes in the dopamine system and within the striatum in heightened reward seeking in adolescence (Ernst et al., 2005; Galván et al., 2006; Van Leijenhorst et al., 2010; Galván and McGlennen, 2013; Chein et al., 2010, Guyer et al., 2009), our data suggest that synaptic remodeling (evidenced by polysialylated neural cell adhesion molecule expression) within the frontal cortex and amygdala may also contribute to enhanced reward sensitivity.

In **Chapter 3**, the data demonstrate that exposure to prescription drugs, methylphenidate and fluoxetine, during the adolescent period of heightened reward sensitivity produces long-term impairments in learning and cognitive flexibility in adulthood in male rats long after drug administration has been terminated. Adolescent methylphenidate exposure has the direst consequences, impairing both the initial learning and reversal of reward contingencies. The data also reveal sex differences: while females take longer to learn the task, they are also less vulnerable to the negative effects of adolescent drug exposure. Analyses of protein expression after learning reveal an upregulation of striatal D1 receptors in adulthood following adolescent

methamphetamine (in males) and methylphenidate (in females) exposure, an upregulation of striatal D2 receptors following adolescent methylphenidate (in females) exposure and higher levels of cortical PSA-NCAM expression in male, compared to female, animals. These findings suggest that developmental psychostimulant exposure may interact with reward experience to boost striatal D1 and D2 receptor expression in a sex-dependent manner later in life. Importantly, these findings have limited implications for learning flexibility and adaptive decision-making in a human clinical population (those diagnosed with attention-deficit/hyperactivity disorder), but may have the most relevance to an adolescent recreational user that consumes psychostimulants as cognitive enhancers to boost academic performance.

In Chapter 5, I present preliminary evidence for enhanced contingent credit assignment (i.e., the attribution of reward feedback to immediately preceding choices) in adolescence: the general pattern of results suggests an enhanced sensitivity to contingent feedback and decreased sensitivity to non-contingent feedback in adolescents compared to adults. While adults integrate information about their previous decisions and past outcomes to guide their subsequent choices, adolescents are more narrowly focused on the most recent choice-reward history. This narrower behavioral focus on recency likely serves a functional purpose as it might better suit the needs of an adolescent in her rapidly changing physical and social environment.

Implications and Future Directions

Benefits and dangers of enhanced adolescent reward sensitivity

In the present study, we extended previous work on adolescent reward sensitivity (van Duijvenvoorde et al., 2016; Doremus-Fitzwater et al., 2010; Steinberg et al., 2009; Urosevic et al., 2012; Andersen et al., 2002) by showing that young rats are willing to invest greater physical effort

to procure rewards, an index of greater sensitivity to positive outcomes. On the one hand, enhanced reward-sensitivity can motivate adolescents to engage in positive and productive behaviors, including academic achievement and the development of passions, hobbies, and strong, positive peer relationships (Telzer, 2016). Adolescents are more efficient than adults, for example, in using extrinsic reward information to improve their own performance (Padmanabhan et al., 2011).

On the other hand, this enhanced sensitivity to natural reinforcers also implies vulnerability to drugs of abuse as drugs hijack the same neural circuitry that evolved to promote survival by motivating food-, water-, shelter-, and mate-seeking behavior. Indeed, adolescent animals have been shown to find psychostimulants such as nicotine (Ahsan et al., 2014; Dannenhoffer and Spear, 2016; Shram et al., 2006; Torres et al., 2008; Levin et al., 2011; Natividad et al., 2013), amphetamine (Shahbazi et al., 2008), methamphetamine (Anker et al., 2012), and cocaine (Brenhouse and Andersen, 2008; Brenhouse et al., 2008; Zakharova et al., 2009; Anker and Carroll, 2010; Wong et al., 2013) more rewarding than their adult counterparts. Adolescent experience with certain drugs, including prescription medication such as methylphenidate and fluoxetine and some illicit drugs depending on the age at the onset of exposure, may in turn endanger subsequent learning and cognitive flexibility in adulthood (Chapter 3; Ye et al., 2014).

The reward-sensitive and reward-centric endophenotype in adolescence has now been demonstrated across species in behaviors including nutrient preference, learning and effortful decision-making (Doremus-Fitzwater and Spear, 2016). Further, the characteristic reward-seeking behaviors of the adolescent period are caused by developmental remodeling in an evolutionary conserved network of brain regions (Doremus-Fitzwater and Spear, 2016). With such degree of cross-species conservation, it is likely that the adolescent reward sensitivity brings a net evolutionary advantage, encouraging environmental exploration, prioritization of rapid acquisition

of nutrients for growth and development, and formation of strong social bonds outside of one's own family. Unlike the brain and behavior, however, our environment has changed tremendously in the past couple of millennia, and it is this rapid change that may have put adolescents in danger, at least with drugs of abuse.

Mammalian brains evolved to forage in environments in which food density varied across geographic locations and through yearly seasons. The long-term evolutionary history in such habitats favored the selection of simple behavioral rules that say that an animals should spend as much time and effort investigating and sampling a particular reward source as that reward source is better than other sources in the environment (Stephens, 2008; Houston and McNamara, 2014; McNamara et al., 2013). In other words, there is no need for an animal to keep foraging in an almost depleted reward source if the alternatives are plentiful. Each reward option is therefore evaluated with respect to the overall reward density in the environment, and during adolescence animals appear to be particularly sensitive to this value difference. Unfortunately, drugs of abuse can hijack this well-calibrated reward-valuation system: the perceived magnitude of the reward brought about by drug exposure is much larger than that following natural reinforcers and that difference might be further magnified for an individual with low environmental density of rewards (Dezfouli et al., 2009). Fortunately, this interpretation also suggests that an increase in environmental enrichment may be protective against drug addiction in adolescence. More specifically, I caution again the approach to policy that prioritizes the removal of reward sources from an adolescent's environment exclusively. Certainly, a complete removal of dangerous reward options (drugs and highly caloric foods) would protect an adolescent from their negative consequences; however, it is extremely unlikely that an exploratory and reward-seeking adolescent would not find a way to sample those rewards. Instead, in my opinion given the data, more

resources should be invested into policies that increase the background level of reward for adolescents through environmental enrichment, exercise, heathy social interaction, education and engagement, because high levels of environmental reward density overall are protective against the influence of any one reward source.

The ugly of recreational cognitive enhancer use in adolescence

The most concerning finding reported in this dissertation is that adolescent exposure to prescription medications, fluoxetine and methylphenidate, has long-lasting consequences for learning and cognitive flexibility in rats. In males, adolescent experience with methylphenidate leads to an impairment both in initial discrimination learning and in later reversal of reward contingencies in adulthood, long after the termination of drug administration. Of utmost importance, these findings apply to recreational use of these substances and have limited implications for the human clinical population. This is because in our experiments, we did not use a rat model of developmental depression or attention-deficit/hyperactivity disorder; instead, we administered methylphenidate and fluoxetine to typical, normally developing rats.

The detrimental effects of methylphenidate on future-life learning and cognitive flexibility are especially concerning, given the recent rise in popularity of cognitive enhancers among human adolescents and young adults (Urban and Gao, 2017; León et al., 2016). Cognitive enhancers are used and abused by college-age students to gain advantage in academic performance and school, work, extracurricular and social productivity by improving focus and wakefulness. Interestingly, in adolescents, methylphenidate use appears to be more similar to that of some illicit drugs in that it is viewed as a rewarding, experimental, 'party' activity, dissociated from work and academic-related pressures (León et al., 2016). From this perspective, increasing the adolescent's estimate

of environmental reward density through enrichment as suggested in the previous section may be beneficial in reducing the propensity for methylphenidate use.

In addition to encouraging positive peer relationships, hobbies and academic achievement, I suggest that aerobic exercise may be a particularly fruitful intervention. Aerobic exercise is perceived as rewarding and has the potential to increase an individual's estimates of the environmental reward density (Thompson et al., 2015; Belke and Wagner, 2005; Kagan and Berkun, 1954). Exercise is protective against addiction at all stages of its progression and can normalize the drug-induced changes in dopamine and glutamate signaling (Lynch et al., 2013; O'Dell et al., 2012; Sobieraj et al., 2014; Engelmann et al., 2014; Smith and Witte, 2012).

Informing pedagogy and education

It has been pointed out before that the accumulating knowledge on adolescent development calls for the abandonment of the Western educational model of one-shot learning in favor of approaches that incorporate trial-by-trial learning with probabilistic contingencies (DePasque and Galván, 2017). I suggest that for most efficient learning, adolescents might also require immediate feedback on their performance. Adolescents are more narrowly focused on the most recent choice-reward history and discount future rewards at a higher rate than adults (Huang et al., 2017). When feedback on assessments is not expected to be timely, adolescents might not be motivated to perform well due to the increase in delay discounting. Furthermore, when unexpected feedback is provided on an assessment long after it was submitted, adolescents may learn suboptimally from that feedback. Finally, as adolescents are still learning-to-learn, they may benefit from frequent practice with gradual, structured introduction to complex problems. This would require a shift in emphasis from summative to formative assessments. While many of these recommendations might

be difficult to implement in the traditional classroom, novel online and gaming solutions (Girard et al., 2013; Ashinoff, 2014) may be explored for their utility as additions to the school environment.

References

Ahsan, H. M., de la Pena, J. B., Botanas, C. J., Kim, H. J., Yu, G. Y., Cheong, J. H. (2014). Conditioned place preference and self-administration induced by nicotine in adolescent and adult rats. *Biomol Ther (Seoul)* 22, 460–466.

Andersen, S. L., Thompson, A. P., Krenzel, E., Teicher, M. H. (2002). Pubertal changes in gonadal hormones do not underlie adolescent dopamine receptor overproduction. *Psychoneuroendocrinology* 27, 683–691.

Anker, J. J., Baron, T. R., Zlebnik, N. E., Carroll, M. E. (2012). Escalation of methamphetamine self-administration in adolescent and adult rats. *Drug Alcohol Depend* 124, 149–153.

Anker, J. J., Carroll, M. E. (2010). Reinstatement of cocaine seeking induced by drugs, cues, and stress in adolescent and adult rats. *Psychopharmacology (Berl)* 208, 211–222.

Ashinoff, B. K. (2014). The potential of video games as a pedagogical tool. Front Psychol. 5, 1109.

Belke, T. W., Wagner, J. P. (2005). The reinforcing property and the rewarding aftereffect of wheel running in rats: a combination of two paradigms. *Behav Process* 68(2), 165–172.

Brenhouse, H. C., Andersen, S. L. (2008). Delayed extinction and stronger reinstatement of cocaine conditioned place preference in adolescent rats, compared to adults. *Behav Neurosci* 122, 460–465.

Brenhouse, H. C., Sonntag, K. C., Andersen, S. L. (2008). Transient D1 dopamine receptor expression on prefrontal cortex projection neurons: relationship to enhanced motivational salience of drug cues in adolescence. *J Neurosci* 28, 2375–2382.

Chein, J., Albert, D., O'Brien, L., Uckert, K., Steinberg, L. (2010). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Dev Sci* 14(2), F1–F10.

Dannenhoffer, C. A., Spear, L. P. (2016). Age differences in conditioned place preferences and taste aversions to nicotine. *Dev Psychobiol* 58, 660–666.

DePasque, S., Galván, A. (2017). Frontostriatal development and probabilistic reinforcement learning during adolescence. *Neurobiol Learn Mem* 143, 1-7.

Dezfouli, A., Piray, P., Keramati, M. M., Ekhtiari, H., Lucas, C., Mokri, A. (2009). A neurocomputational model for cocaine addiction. *Neural Comput* 21(10), 2869-93.

Doremus-Fitzwater, T. L., Spear, L. P. (2016). Reward-centricity and attenuated aversions: An adolescent phenotype emerging from studies in laboratory animals. *Neurosci Biobehav Rev* 70, 121–134.

Doremus-Fitzwater, T. L., Varlinskaya, E. I., Spear, L. P. (2010). Motivational systems in adolescence: possible implications for age differences in substance abuse and other risk-taking behaviors. *Brain Cogn* 72, 114–123.

Engelmann, A. J., Aparicio, M. B., Kim, A., Sobieraj, J. C., Yuan, C. J., Grant, Y., et al. (2014). Chronic wheel running reduces maladaptive patterns of methamphetamine intake: regulation by attenuation of methamphetamine-induced neuronal nitric oxide synthase. *Brain Struct Funct* 219(2), 657–672.

Ernst, M., Nelson, E. E., Jazbec, S., McClure, E. B., Monk, C. S., Leibenluft, E., et al. (2005). Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. *Neuroimage* 25(4), 1279–1291.

Friemel, C. M., Spanagel, R., Schneider, M. (2010). Reward sensitivity for a palatable food reward peaks during pubertal developmental in rats. *Front Behav Neurosci* 2010;4

Galván, A., McGlennen, K. (2013). Enhanced striatal sensitivity to aversive reinforcement in adolescents versus adults. *J Cogn Neurosci* 25, 284–296.

Galván, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., et al. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *J Neurosci* 26(25), 6885–6892.

Girard, C., Ecalle, J., Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *J Comput Assist Learn* 29, 207–219.

Guyer, A. E., McClure-Tone, E. B., Shiffrin, N. D., Pine, D. S., Nelson, E. E. (2009). Probing the neural correlates of anticipated peer evaluation in adolescence. *Child Dev* 80(4), 1000–1015.

Houston, A. I., McNamara, J. M. (2014). Foraging currencies, metabolism and behavioural routines. *J Anim Ecol.* 83(1), 30-40.

Huang, Y., Hu, P., Li, X. (2017). Undervaluing delayed rewards explains adolescents' impulsivity in inter-temporal choice: an ERP study. *Sci Rep* 7, 42631.

Kagan, J., Berkun, M. (1954). The reward value of running activity. *J Comp Physiol Psychol* 47(2), 108.

León, K. S., Martínez, D. E. (2017). To Study, to Party, or Both? Assessing Risk Factors for Non-Prescribed Stimulant Use among Middle and High School Students. *J Psychoactive Drugs* 49(1), 22-30.

Levin, E. D., Slade, S., Wells, C., Cauley, M., Petro, A., Vendittelli, A., et al. (2011). Threshold of adulthood for the onset of nicotine self-administration in male and female rats. *Behav Brain Res* 225, 473–481.

Lynch, W. J., Peterson, A. B., Sanchez, V., Abel, J., Smith, M. A. (2013). Exercise as a novel treatment for drug addiction: a neurobiological and stage-dependent hypothesis. *Neurosci Biobehav Rev* 37(8), 1622–1644.

McNamara, J. M., Fawcett, T. W., Houston, A. I. (2013). An adaptive response to uncertainty generates positive and negative contrast effects. *Science* 340(6136), 1084-6.

Natividad, L. A., Torres, O. V., Friedman, T. C., O'Dell, L. E. (2013). Adolescence is a period of development characterized by short- and long-term vulnerability to the rewarding effects of nicotine and reduced sensitivity to the anorectic effects of this drug. *Behav Brain Res* 257, 275–285.

O'Dell, S. J., Galvez, B. A., Ball, A. J., Marshall, J. F. (2012). Running wheel exercise ameliorates methamphetamine-induced damage to dopamine and serotonin terminals. *Synapse* 66(1), 71–80.

Padmanabhan, A., Geier, C. F., Ordaz, S. J., Teslovich, T., Luna, B. (2011). Developmental changes in brain function underlying the influence of reward processing on inhibitory control. *Dev Cogn Neurosci* 1(4), 517–529.

Shahbazi, M., Moffett, A. M., Williams, B. F., Frantz, K. J. (2008). Age- and sex-dependent amphetamine self-administration in rats. *Psychopharmacology (Berl)* 196, 71–81.

Shram, M. J., Funk, D., Li, Z., Le, A. D. (2006). Periadolescent and adult rats respond differently in tests measuring the rewarding and aversive effects of nicotine. *Psychopharmacology (Berl)* 186, 201–208.

Smith, M. A., Witte, M. A. (2012). The effects of exercise on cocaine self-administration, food-maintained responding, and locomotor activity in female rats: importance of the temporal relationship between physical activity and initial drug exposure. *Exp Clin Psychopharmacol* 20(6), 437–446.

Sobieraj, J. C., Kim, A., Fannon, M. J., Mandyam, C. D. (2014). Chronic wheel running-induced reduction of extinction and reinstatement of methamphetamine seeking in methamphetamine dependent rats is associated with reduced number of periaqueductal gray dopamine neurons. *Brain Struct Funct* 221(1): 261–276.

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., Banich, M. (2009). Age differences in future orientation and delay discounting. Child Dev. 80, 28–44.

Stephens, D. W. (2008). Decision ecology: foraging and the ecology of animal decision making. Cogn *Affect Behav Neurosci.* 8(4), 475-84.

Sullivan-Toole, H., DePasque, S., Holt-Gosselin, B., Galván, A. (2019). Worth working for: The influence of effort costs on teens' choices during a novel decision making game. *Dev Cogn Neurosci.* 37, 100652.

Telzer, E. H. (2016). Dopaminergic reward sensitivity can promote adolescent health: A new perspective on the mechanism of ventral striatum activation. *Dev Cogn Neurosci.* 17, 57-67.

Thompson, A. B., Stolyarova, A., Ying, Z., Zhuang, Y., Gómez-Pinilla, F., Izquierdo, A. (2015). Methamphetamine blocks exercise effects on Bdnf and Drd2 gene expression in frontal cortex and striatum. *Neuropharmacology* 99, 658-64.

Torres, O. V., Tejeda, H. A., Natividad, L. A., O'Dell, L. E. (2008). Enhanced vulnerability to the rewarding effects of nicotine during the adolescent period of development. *Pharmacol Biochem Behav* 90, 658–663.

Urban, K. R., Gao, W. J. (2017). Psychostimulants As Cognitive Enhancers in Adolescents: More Risk than Reward? 5, 260.

Urosevic, S., Collins, P., Muetzel, R., Lim, K., Luciana, M. (2012). Longitudinal changes in behavioral approach system sensitivity and brain structures involved in reward processing during adolescence. *Dev Psychol* 48, 1488–1500.

van Duijvenvoorde, A. C., Peters, S., Braams, B. R., Crone, E. A. (2016). What motivates adolescents? Neural responses to rewards and their influence on adolescents' risk taking, learning, and cognitive control. *Neurosci Biobehav Rev* 70, 135-147.

Van Leijenhorst, L., Moor, B. G., de Macks, Z. A. O., Rombouts, S. A., Westenberg, P. M., Crone, E. A. (2010). Adolescent risky decision-making: neurocognitive development of reward and control regions. *Neuroimage* 51(1), 345–355.

Wong, W. C., Ford, K. A., Pagels, N. E., McCutcheon, J. E., Marinelli, M. (2013). Adolescents are more vulnerable to cocaine addiction: behavioral and electrophysiological evidence. *J Neurosci* 33, 4913–4922.

Ye, T., Pozos, H., Phillips, T. J., Izquierdo, A. (2014). Long-term effects of exposure to methamphetamine in adolescent rats. *Drug Alcohol Depend* 138, 17–23.

Zakharova, E., Wade, D., Izenwasser, S. (2009). Sensitivity to cocaine conditioned reward depends on sex and age. *Pharmacol Biochem Behav*. 92, 131–134.