Summer 8-15-2015

Breaking Apart the Reinforcement Learning Deficit in Schizophrenia

Adam Culbreth
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychology

Breaking Apart the Reinforcement Learning Deficit in Schizophrenia

by
Adam J. Culbreth

A thesis presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the
degree of Master of Arts

August 2015

St. Louis, Missouri

Table of Contents

List of Figures iii

List of Tables iv

Acknowledgments v

Abstract vi

Introduction 1

Methods 4

Results 6

Discussion 9

References 17

Tables 20

Figures 24

List of Figures

Figure 1. Sample trial diagram & Task Structure

Figure 2. First-stage choice behavior (coded as stay/shift) averaged across individuals within each group.

List of Tables

Table 1. Participant Characteristics

Table 2. Coefficients predicting response repetition from the outcome of the previous trial, the transition type, and diagnostic group.

Table 3. External Correlates of model-based and model-free learning

Table 4. Second stage choice reaction time by group and transition type.

Acknowledgements

I would like to acknowledge Dr. Deanna Barch for helpful mentorship throughout this study. I would also like to thank our collaborators, Andrew Westbrook, Dr. Nathaniel Daw (New York University), Dr. Matthew Botvinick (Princeton University). I also thank Catherine Hartley for providing the modified task stimuli utilized in this study. This work was supported by National Institute of Mental Health R01 MH066031. Finally, I would like to acknowledge the participants of the current research study without whom any of this work could have been completed.

ABSTRACT OF THE THESIS

Breaking Apart the Reinforcement Learning Deficit in Schizophrenia

by

Adam J. Culbreth

Master of Arts in Psychology

Washington University in St. Louis, 2015

Professor Deanna M. Barch, Chair

Reinforcement learning deficits have long been associated with schizophrenia. However, tasks traditionally used to assess these deficits often rely on multiple processing streams leaving the etiology of these task deficits unclear. In the current study, we borrowed a recent framework from computational neuroscience, which separates reinforcement-learning into two distinct systems, model-based and model-free. Under this framework, the model-free system learns about the value of actions in the immediate context, while the model-based system learns about the value of actions in both immediate and subsequent states that may be encountered as a result of their actions. Using a decision task that has been previously validated to assess relative reliance on each system we showed that individuals with schizophrenia demonstrated decreased model-based but intact model-free learning estimates. Furthermore, parameter estimates of model-based behavior correlated positively with IQ, suggesting that model-based deficits in schizophrenia may relate to reduced intellectual function. These findings specify reinforcement-learning deficits in schizophrenia by showing both intact and disturbed components. Such findings and computational frameworks provide meaningful insights as researchers continue to characterize decision-making circuitry in schizophrenia as a means to discover new pathways for interventions.

Introduction

Deficits in various aspects of reinforcement learning (RL) have long been associated with schizophrenia (SZ) (Barch & Dowd, 2010; Gold, Waltz, Prentice, Morris, & Heerey, 2008). Importantly, such deficits have shown relationships to other core features of the disorder, such as negative symptom severity (Strauss et al., 2011; Waltz, Frank, Robinson, & Gold, 2007). However, tasks aimed at assessing these deficits often rely on multiple processing streams (e.g., reward processing, cognitive control, attention, and working memory) leaving the specific causes of these task deficits unclear. Thus, one concern regarding RL deficits in SZ lies in delineating intact from disturbed processing components. Such findings will be imperative to not only our etiological understanding of SZ, but also when considering treatment approaches, as deficits in different processing streams will likely lend themselves to differentially effective interventions. The current study uses a recent framework from computational neuroscience to separate RL processes. This framework posits that decision-making behavior arises from two parallel control systems, model-free (also thought of as habitual control) and model-based (also thought of as goal-directed control). The model-free system learns, retrospectively through repeating previously rewarded actions, while the model-based system learns prospectively by weighing future outcomes based on an "internal model" of the environment. Using this framework, the current study aims to examine the relative contributions of deficits in model-free and model-based RL, as a means of specifying the RL deficit in SZ (Daw, Niv, & Dayan, 2005).

As referred to above, current theorists suggest that individuals utilize two systems, commonly referred to as habitual (model-free) and goal-directed (model-based) control, to guide decision-making (Balleine, Daw, & O'Doherty, 2008; Dickinson & Balleine, 2002). In habitual control, individuals make decisions by repeating previously rewarded actions (or not repeating punished actions) (Thorndike, 1933). For example, children learn not to touch a hot stove by getting burned previously. In contrast, goal-directed control requires individuals to make decisions by prospectively weighing future outcomes based on an "internal model" of the environment (Tolman, 1948). For example, expert chess players make

decisions by thinking about the future state of the chessboard. Recently, this decision-making model has been placed into the computational framework of model-based and model-free RL in order to more quantitatively assess the individual contribution, integration, and structure of these systems (Balleine et al., 2008; Daw et al., 2005). In model-free RL, the actor learns preferences for actions directly by repeating reinforced choices, without taking into consideration a 'model' of the environment. This system is driven largely by reward prediction error and the ventral striatum (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Gläscher, Daw, Dayan, & O'Doherty, 2010). In contrast, model-based RL requires the actor to learn an internal representation of the structure of the environment and utilize this structure to make decisions. The neural correlates of this system include lateral prefrontal cortex (Gläscher et al., 2010), dorsolateral prefrontal cortex (dlPFC) (Smittenaar, FitzGerald, Romei, Wright, & Dolan, 2013), orbitofrontal cortex (OFC) (Lee, Shimojo, & O'Doherty, 2014; McDannald, Lucantonio, Burke, Niv, & Schoenbaum, 2011), ventromedial prefrontal cortex (vmPFC) (Gläscher et al., 2010; Lee et al., 2014), and striatal regions (Daw et al., 2011; McDannald et al., 2011). Recent work has also pointed to a critical role of ventral striatal presynaptic dopamine in potentially modulating the relative contributions of these control systems (Deserno et al., 2015).

Previous reports have explored processes associated with both habitual and goal-directed control in SZ. For example, individuals with SZ have impairments in processes associated with model-based RL such as proactive cognitive control and working memory capacity (Barch & Ceaser, 2012; Otto, Raio, Chiang, Phelps, & Daw, 2013; Otto, Skatova, Madlon-Kay, & Daw, 2014). SZ patients also have functional and structural abnormalities in brain regions thought to be associated with model-based RL (dlPFC: (Barch, Csernansky, Conturo, & Snyder, 2002; Minzenberg, Laird, Thelen, Carter, & Glahn, 2009; Semkovska, Bédard, Godbout, Limoge, & Stip, 2004); vmPFC: (Hooker, Bruce, Lincoln, Fisher, & Vinogradov, 2011; Park, Park, Chun, Kim, & Kim, 2008); OFC: (Gur et al., 2000; Pantelis et al., 2003; Plailly, d'Amato, Saoud, & Royet, 2006). In contrast, individuals with SZ often show intact functioning in processes associated with model-free RL, such as, procedural learning (Kéri et al., 2000; Weickert et al.,

2002), implicit RL (Heerey, Bell-Warren, & Gold, 2008), and intact striatal responses to reward receipt (Dowd & Barch, 2012; Simon et al., 2010; Waltz et al., 2010). These findings are also consistent with a recent RL study that utilized a computational model to reveal that working memory impairments made a significant contribution to the RL deficits in SZ with little evidence for alterations in the basic stimulus-response component of RL (Collins, Brown, Gold, Waltz, & Frank, 2014). Neuroimaging results from our group have also found that patient deficits in a cognitive control network of brain regions predicted task behavior on a RL task, whereas, basic reward processing regions did not, which is consistent with the hypothesis of a model-based RL deficit in SZ (Culbreth et al., in review). However, despite this converging evidence, the computational framework of model-based and model-free RL has not been explicitly tested in SZ. Assessing these deficits from a computational standpoint may allow for better characterization of RL dysfunction in SZ, as well as, provide testable hypotheses about systems to target for novel interventions (Montague, Dolan, Friston, & Dayan, 2012)

The current study aimed to use this computational and experimental framework to examine the relative integrity of model-free versus model-based RL in SZ. We utilized a 2-stage Markov decision-making task created by Daw, which has been previously validated to separately assess model-based and model-free choice behavior (See Figure 1) (Daw et al., 2011; Otto et al., 2013; Otto et al., 2014; Smittenaar et al., 2013; Voon et al., 2014; Wunderlich, Smittenaar, & Dolan, 2012). In this task, individuals make a 2-alternative forced choice, which leads to a common (70% of the time) or rare second-stage, where a second 2-alternative forced choice is made in hopes of receiving a reward. The reward contingencies of choices in the second-stage fluctuate, and participants adopt different strategies to maximize rewarded outcomes on the task. If individuals make first-stage choices based solely on the previous trial being rewarded they are considered model-free (i.e., habitual). However, if they assess the rarity of the transition along with the prior rewards (i.e. build an internal model of the task environment) they are considered model-based (i.e., goal-directed). For example, model-based RL would predict that individuals would be more likely to shift their initial response following a previously rewarded outcome

with a rare transition.

Utilizing this task, we hypothesized that individuals with SZ would show intact model-free learning estimates. This prediction is consistent with previous reports, which show intact procedural/habit learning in SZ (Kéri et al., 2000; Weickert et al., 2002), as well as, intact striatal responses to reward receipt (Dowd & Barch, 2012; Simon et al., 2010; Waltz et al., 2010). We hypothesized that individuals with SZ would show impaired model-based learning estimates on the Markov decision-making task. This prediction is consistent with literature suggesting that SZ patients have deficits in proactive cognitive control and working memory capacity (Barch & Ceaser, 2012), as well as, functional and structural abnormalities in brain regions associated with model-based RL (e.g. dlPFC). Furthermore, we hypothesized that these learning estimates would be correlated with negative symptoms and measures of neurocognitive functioning (i.e., full-scale IQ), such that, individuals with greater negative symptom severity and lower neurocognitive functioning will demonstrate decreased model-based learning estimates.

Methods

Participants

Participants were 33 individuals meeting DSM-IV criteria for SZ or schizoaffective disorder (SZA; N=13), and 30 controls (CN), with no personal or family history of psychosis, from the Saint Louis community. Seven SZ patients were unmedicated. Exclusion criteria included 1) DSM-IV diagnosis of substance abuse or dependence in the past six months; 2) DSM-IV diagnosis of major depressive disorder or dysthymia in the past year; 3) changes in medication dosage two weeks prior to consent; 4) past head injury with documented neurological sequelae and/or loss of consciousness; 5) mental retardation. The Washington University Institutional Review Board approved the study. Participants provided written informed consent in accordance with Washington University's Human Subject Committee's criteria.

Clinical Assessments

Diagnoses were determined by the Structured Clinical Interview for DSM-IV-TR (First, Spitzer, Gibbon, & Williams, 2001). Negative Symptoms were assessed using the Brief Negative Symptom Scale

(BNSS) (Kirkpatrick et al., 2011). Avolition and anhedonia were also assessed using the following self-report measures: The Revised Chapman Physical and Social Anhedonia Scales (Chapman, Chapman, & Raulin, 1976, 1978) and the Snaith-Hamilton Pleasure Scale (Snaith et al., 1995). All participants we required to pass a urine drug screen and a breathalyzer test.

Task

Participants completed a modified version of a two-stage learning task developed by Daw and colleagues (See Figure 1 for sample trial) (Daw et al., 2011). Extensive instructions on the task structure and practice trials were completed prior to task administration. At the start of each trial, a two-alternative forced choice was presented. This choice then led, probabilistically, to one of two, second-stage "states" (Figure 1). Importantly, each choice led more frequently (70%) to one of two "states". During the second-stage, another two-alternative forced choice was presented and feedback was received (i.e., reward or no reward). In order to ensure learning throughout the task, the probability of receiving a reward for the four, second-stage options changed slowly using Gaussian random walks. All together, participants completed 200 trials, and had 3 seconds to make a response for each choice. Interstimulus and intertrial intervals were 1 second. Reward outcome was presented for 1 second. Stimuli on the task consisted of spaceships and aliens, instead of the Tibetan characters in the original Daw paper(Daw et al., 2011), to provide a more interactive task environment for better compliance and understanding in our patient population.

Data Analysis

Logistic Regression

We conducted a mixed-effects logistic regression following the protocol of Daw and colleagues (Daw et al., 2011). Analyses were performed using the lme4 linear mixed effects package in R (Bates & Sarkar, 2007). The dependent variable was the first stage choice (coded as stay or shift). Explanatory variables were whether the previous trial was rewarded or not rewarded, whether the previous trial's transition was rare or common, and the interaction between rarity and reward. All coefficients were taken as random effects across participants, and estimates are reported at the population level. In order to

observe group differences between model-based and model-free learning estimates diagnostic group was coded as a binary variable and interacted as a factor. Finally, full scale IQ as determined by the WTAR, BNSS negative symptoms scores, and the Snaith-Hamilton pleasure scale were group-centered and interacted as covariates to determine individual difference relationships.

Reaction Time

In order to observe the relative contributions of model-based and model-free control, we also analyzed reaction time at stage 2. We conducted a repeated-measures ANOVA with 2 factors, diagnostic group and transition type. In order to observe individual difference relationships of second-stage reaction time differences between transition types, partial correlations were conducted between an RT difference score (rare – common) and BNSS rated negative symptoms, full-scale IQ, and the Snath-Hamilton pleasure scale.

Results

Demographics

The groups did not significantly differ in age, gender, ethnicity, or parental education (Table 1). However, as expected, the personal education of the SZ group was lower than the CN group. The SZ group self-reported increased levels of anhedonia compared to the CN group across all measures (Table 1). Table 1 also describes the current prescribed medications of the SZ group.

Task Behavior

In order to assess the relative contributions of model-based and model-free learning we analyzed first-stage choice behavior (coded as stay or shift). As a reminder, model-based and model-free strategies make different predictions about how previous second-stage rewards should influence current first-stage choices. For example, consider a trial where a first-stage choice results in a rare transition leading to an uncharacteristic second-stage state where a reward is obtained (rare rewarded condition in Figure 2). The pure model-free learner would repeat the first-stage choice on the following trial, basing decision-making solely on reward and ignoring the transition structure (Supplemental Material 1). However, the model-

based learner, taking into consideration both reward and task structure, would show a decreased likelihood of repeating the same first-stage choice because shifting to the other first-stage choice would lead with greater probability to the previously rewarded second-stage state (Supplemental Material 1). Figure 2 illustrates the frequency of staying with the previous first-stage choice for the SZ and CN groups. Qualitatively, the SZ group choice behavior appears less model-based and more model-free than the CN behavior (Supplemental Materials 1).

1. Hierarchical Linear Model

In order to quantitatively analyze choice behavior, we then conducted a mixed-effects logistic regression, which predicted the current first-stage choice (coded as shift or stay) as a function of the previous trial's outcome (reward), the previous transition (rarity), and diagnostic status (group) (Table 2). There was a significant effect of reward, such that the probability of staying with the same first-stage choice increased when the previous choice was rewarded, demonstrating model-free learning. There was also a significant reward x rarity interaction, such that participants were less likely to repeat the prior first-stage choice that led to a reward if it was a preceded by a rare versus a more common transition (see Table 2). Furthermore, there was a significant reward x rarity x group interaction, such that the SZ group demonstrated less evidence of model-based learning compared to the CN group (see Table 2 and further analyses below). Finally, although qualitatively greater in the SZ group, the group x reward interaction failed to reach significance indicating that the CN and SZ groups did not differ on estimates of model-free learning.

In order to further examine the source of the significant reward x rarity x group interaction we conducted separate HLM analyses for the CN and SZ groups (Supplemental Materials 1). In the CN analysis, participants showed significant and robust effects of both model-based (rarity X reward) and model-free learning (reward) (Supplemental Materials 1). However, in the SZ group, participants showed a significant and robust effect of model-free learning (reward) while only showing a trend level model-

based estimate (rarity X reward) (Supplemental Materials 1). These within-group comparisons strengthen our conclusion from the reward X rarity X group interaction that model-based learning seems to be reduced in the SZ group (Supplemental Materials 1).

Follow-up analyses were conducted to determine individual difference metrics that may account for increased/decreased, model-based/model-free learning estimates (Table 3). Contrary to our hypotheses, learning estimates did not show a significant relationship with clinician-rated negative symptoms (BNSS total score). However, model-based learning estimates did show a significant interaction with full-scale IQ while controlling for diagnostic status, suggesting that individuals with greater intellectual capacity utilized model-based control more than those with reduced intellectual functioning (i.e., a more positive beta weight for reward X rarity interaction) (see Table 3). Importantly, this association between model-based learning and IQ was found separately in both the SZ and the CN groups. Finally, model-free learning estimates showed a significant interaction with the total score of the Snaith-Hamiliton Pleasure Scale, indicating that participants with higher levels of hedonic capacity demonstrated a greater effect of reward (i.e., a larger reward effect) (Table 3). However, this relationship was not significant in the CN group.

2. Reaction Time

As a second means of assessing model-based and model-free control, we analyzed second-stage reaction time. Importantly, model-free RL would not predict differential second-stage reaction times following rare and common transitions, as this form of learning carries no information of state transitions (Deserno et al., 2015). Thus, any differences in the second-stage reaction-time between transition states should be due to model-based learning. Using a repeated-measures ANOVA, we found a significant main effect of transition on RT, where choices following a rare transition were slower than those following common transitions ($F=38.60$, $p<0.001$) (see Table 4 for descriptives). Contrary to our hypotheses, the interaction between transition type and diagnostic group was not significant ($F=1.64$, $p=0.2$). However, when analyzing the simple effect of RT within each group separately, we found that the size of the

transition effect for the CN group ($\eta^2_{partial} = 0.305$) was almost twice the size of the SZ group ($\eta^2_{partial} = 0.173$) suggesting a more robust level of involvement of model-based RL for the CN group. Finally, a significant partial correlation was found between RT difference (rare-common) and IQ holding diagnostic group constant ($r = 0.301$) providing converging evidence for the aforementioned relationship between greater IQ and greater model-based RL estimates. However, no significant relationship was found between RT and self-reported or clinician-rated negative symptoms.

Discussion

The goal of the current experiment was to assess the relative contributions of model-based and model-free learning in SZ. Consistent with our hypotheses, individuals with SZ demonstrated reduced model-based learning estimates compared to healthy controls, suggesting a deficit in their ability to create and utilize an "internal model" of the task environment to drive goal-oriented decision-making. In contrast, individuals with SZ did not differ from healthy controls in regards to model-free learning estimates, suggesting intact habitual learning. Supplementary analyses of second-stage reaction time indicated evidence for model-based learning in both groups. In general, participants took longer to respond for second stage choices preceded by a rare transition compared to a common transition. However, the size of this effect was almost twice as large in the patients compared to the controls, suggesting that while the groups did not significantly differ in regards to reaction time controls showed a much more robust effect of our model-based estimate. In regards to individual differences, model-based learning estimates were positively correlated to measures of intellectual functioning in both the patient and the control group. However, the hypothesis that model-based and model-free learning estimates would be related to negative symptom severity was not supported.

The current results are consistent with several previous reports related to cognitive and affective dysfunction in SZ. The results of reduced model-based estimates are consistent with previous reports showing that individuals with SZ have deficits in processing domains frequently associated with model-based RL, such as proactive cognitive control and working memory capacity (Barch & Ceaser, 2012). The

results of intact model-free estimates are consistent with results showing intact procedural learning and implicit RL in SZ (Heerey et al., 2008; Kéri et al., 2000; Weickert et al., 2002). Furthermore, these results are also conceptually consistent with a recent report by Collins and colleagues who used a computational model to illustrate that working-memory impairments made a significant contribution to impairments on a RL task in SZ, with little evidence for alternations in the basic stimulus-response component of RL. Interestingly, Collins and colleagues also failed to show a relationship of task deficits and negative symptoms (Collins et al., 2014). This converging evidence suggests that perhaps the RL deficit seen in SZ relates more to levels of higher-order cognitive functioning than negative symptoms.

The current results are also consistent with several previous reports examining the external correlates of model-based and model-free learning. As stated in the introduction, increased model-based behavior has been shown to be positively correlated with higher-order cognitive processes, such as proactive cognitive control and working memory capacity (Otto et al., 2013; Otto et al., 2014). Consistent with these findings, we found a significant relationship between model-based learning and IQ, suggesting that participants with higher levels of intellectual functioning were more likely to use model-based strategies. This relationship was found in both the SZ and the CN groups. Interestingly, we also found a significant relationship between a measure of hedonic capacity (Snaith total score) and model-free learning, such that individuals with higher hedonic capacity illustrated a greater reward effect. These external correlates provide further evidence for and characterize the separable systems, which underlie model-based and model-free learning by suggesting that these systems may be associated with differential external factors (i.e., reward processing and basic hedonics for model-free and higher-order cognition for model-based). However, the current study is unable to disentangle the causal direction of these relationships. For instance in regards to IQ, one unanswered question is do individuals with SZ have model-based deficits because they also tend to have lower IQ estimates, or are decreased IQ estimates and reduced model-based learning both manifestations of a common neural dysfunction? It will be important for future reports to examine individual difference relationships with larger sample sizes in order to

elucidate the nature of the relationship between these factors and how they unfold over time.

The current study yields an important finding regarding RL deficits in SZ. However, we believe that a further benefit of the current approach lies in the directly testable hypotheses about dysfunctional neural systems that emerge from our initial result of impaired model-based RL in SZ and more broadly from the model-based/model-free framework. Importantly, this framework provides accounts of the biological systems that underlie these different forms of RL. Now that we have established evidence suggesting a model-based deficit in SZ, future studies can leverage this framework to discern which neural mechanisms, within the model-based system, are associated with this deficit. Previous literature suggests three likely biological targets for such a deficit. 1) abnormal presynaptic dopamine in the ventral striatum (Deserno et al., 2015); 2) abnormal functioning in the lateral prefrontal cortex (Deserno et al., 2015; Gläscher et al., 2010); 3) abnormal cortico-striatal connectivity. Importantly, as noted in the introduction, individuals with SZ have shown functional deficits related to all of these possible biological targets. Such studies will be critical as they will allow researchers to specify the neural structure of the RL deficit in SZ, and will possibly result in the generation of biologically-informed targets for novel interventions.

The current findings also have implications for current intervention strategies aimed at improving RL in SZ. Our findings suggest that the RL deficit in SZ results largely from deficits in model-based processing. Thus, these initial findings suggest cognitive-remediation interventions aimed at improving cognitive domains frequently associated with model-based learning (e.g., cognitive control) might be beneficial at alleviating the RL deficits associated with SZ. Indeed, preliminary evidence suggests that cognitive remediation may be beneficial for improving performance on a Wisconsin Card Sorting Task (Cella et al., 2014), which taps into at least some of the processes relevant for model-based learning. However, it is unknown whether these improvements translate to other RL tasks. Future research needs to be conducted to discern whether these cognitive-remediation interventions might be beneficial at restoring model-based learning in individuals with SZ.

Limitations

The current study has several limitations. First, although we saw robust effects of both model-based and model-free learning estimates in the CN group we also saw a main effect of transition type, where participants had a tendency to shift first-stage responses more for trials following a common compared to a rare transition. This finding is inconsistent with previous reports, and is particularly perplexing being as neither model-based nor model-free RL would predict an effect of transition (Daw et al., 2011; Deserno et al., 2015; Otto et al., 2013; Otto et al., 2014). However, qualitatively our data appears consistent with previous reports suggesting that while this effect was unexpected it likely did not influence our model-based and model-free estimates (Daw et al., 2011) (Supplemental Materials 1). Second, the sample size of the current study may not have been sufficient to comprehensively examine the relationship of individual difference factors to learning estimates. However, we were able to show individual difference relationships between intellectual functioning and model-based RL. Future studies will need to be conducted with larger sample sizes and more diverse individual difference metrics in order to better characterize relationships between learning estimates and external factors. Third, in the current study we utilized a hierarchical linear model to analyze choice behavior on our sequential learning task. While the HLM approach has been previously validated to robustly quantify model-based and model-free learning more sophisticated computational models may allow for further separation of model-based and model-free components between groups providing useful information for guiding future directions of research (Daw et al., 2011(Gläscher, 2010 #17). Such models also allow us to further characterize the model-based learning deficit in SZ by assessing key variables such as learning rate, which will be beneficial in assessing whether individuals with SZ are slow or unable to build "internal models" of their environment. Thus, follow-up analyses using more detailed computational models will be conducted in order to glean additional information possibly masked by the HLM approach. Finally, the majority of our SZ sample was currently taking anti-psychotic medications, which may have altered reward-related responses.

Summary

The goal of the current study was to utilize a recent computational framework to separately assess the relatively contribution of RL processing streams associated with SZ. We found evidence for reduced model-based learning estimates in individuals with SZ compared to healthy controls, suggesting that individuals with SZ have difficulties creating and utilizing an "internal model" of their environment to drive goal-oriented decision-making. However, model-free learning estimates were intact in individuals with SZ, suggesting that basic habitual RL processing appears unaffected. This finding is in line with multiple reports that show that individuals with SZ have deficits in processing domains that are commonly associated with model-based learning such as cognitive control and working memory capacity. Importantly, the current framework allows for the testing of distinct hypotheses about the nature of RL deficit in SZ. Future studies will be needed to better elucidate deficits in neural circuitry, which might cause the behavioral abnormalities seen in the current study. For example, is the model-based learning deficit a result of poor striatal functioning, prefrontal functioning, or aspects of cortico-striatal connectivity. Such investigations will aid in understanding the etiology of this deficit and provide biological targets for further inquiry as researchers attempt to better characterize decision-making neural circuitry in order to discover new pathways for interventions.

References

Balleine, B. W., Daw, N. D., & O'Doherty, J. P. (2008). Multiple forms of value learning and the function of dopamine. *Neuroeconomics: decision making and the brain*, 367-385.

Barch, D. M., & Ceaser, A. (2012). Cognition in schizophrenia: core psychological and neural mechanisms. *Trends in cognitive sciences, 16*(1), 27-34.

Barch, D. M., Csernansky, J. G., Conturo, T., & Snyder, A. Z. (2002). Working and long-term memory deficits in schizophrenia: is there a common prefrontal mechanism? *Journal of abnormal psychology, 111*(3), 478.

Barch, D. M., & Dowd, E. C. (2010). Goal representations and motivational drive in schizophrenia: the role of prefrontal–striatal interactions. *Schizophrenia Bulletin, 36*(5), 919-934.

Bates, D., & Sarkar, D. (2007). Ime4: Linear mixed-effects models using S4 classes.'R'package. Version 0.9975-12. *URL http://CRAN. R-project. org*.

Cella, M., Bishara, A. J., Medin, E., Swan, S., Reeder, C., & Wykes, T. (2014). Identifying Cognitive Remediation Change Through Computational Modelling—Effects on Reinforcement Learning in Schizophrenia. *Schizophrenia bulletin, 40*(6), 1422-1432.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1976). Scales for physical and social anhedonia. *Journal of abnormal psychology, 85*(4), 374.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1978). Body-image aberration in Schizophrenia. *Journal of abnormal psychology, 87*(4), 399.

Collins, A. G. E., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia. *The*

Journal of Neuroscience, 34(41), 13747-13756. doi: 10.1523/jneurosci.0989-14.2014

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

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704-1711.

Deserno, L., Huys, Q. J., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A. A., . . . Schlagenhauf, F. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proceedings of the National Academy of Sciences*, 201417219.

Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. *Stevens' handbook of experimental psychology*.

Dowd, E. C., & Barch, D. M. (2012). Pavlovian reward prediction and receipt in schizophrenia: relationship to anhedonia. *PloS one, 7*(5), e35622.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2001). Structured Clinical Interview for DSM-IV-TR Axis I Disorders—Patient Edition (SCID-I/P. 2/2001 Revision). *New York: Biometrics Research Department, New York State Psychiatric Institute*.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*(4), 585-595.

Gold, J. M., Waltz, J. A., Prentice, K. J., Morris, S. E., & Heerey, E. A. (2008). Reward processing in schizophrenia: a deficit in the representation of value. *Schizophrenia*

bulletin, 34(5), 835-847.

Gur, R. E., Cowell, P. E., Latshaw, A., Turetsky, B. I., Grossman, R. I., Arnold, S. E., . . . Gur, R. C. (2000). Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. *Archives of general psychiatry, 57*(8), 761-768.

Heerey, E. A., Bell-Warren, K. R., & Gold, J. M. (2008). Decision-making impairments in the context of intact reward sensitivity in schizophrenia. *Biological psychiatry, 64*(1), 62-69.

Hooker, C. I., Bruce, L., Lincoln, S. H., Fisher, M., & Vinogradov, S. (2011). Theory of mind skills are related to gray matter volume in the ventromedial prefrontal cortex in schizophrenia. *Biological psychiatry, 70*(12), 1169-1178.

Kéri, S., Kelemen, O., Szekeres, G., Bagoczky, N., Erdelyi, R., Antal, A., . . . Janka, Z. (2000). Schizophrenics know more than they can tell: probabilistic classification learning in schizophrenia. *Psychological medicine, 30*(01), 149-155.

Kirkpatrick, B., Strauss, G. P., Nguyen, L., Fischer, B. A., Daniel, D. G., Cienfuegos, A., & Marder, S. R. (2011). The brief negative symptom scale: psychometric properties. *Schizophrenia bulletin, 37*(2), 300-305.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron, 81*(3), 687-699.

McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *The Journal of Neuroscience, 31*(7), 2700-2705.

Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S., & Glahn, D. C. (2009). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Archives of General Psychiatry, 66*(8), 811-822.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in cognitive sciences, 16*(1), 72-80.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences, 110*(52), 20941-20946.

Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2014). Cognitive Control Predicts Use of Model-based Reinforcement Learning.

Pantelis, C., Velakoulis, D., McGorry, P. D., Wood, S. J., Suckling, J., Phillips, L. J., . . . Soulsby, B. (2003). Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison. *The Lancet, 361*(9354), 281-288.

Park, I. H., Park, H.-J., Chun, J.-W., Kim, E. Y., & Kim, J.-J. (2008). Dysfunctional modulation of emotional interference in the medial prefrontal cortex in patients with schizophrenia. *Neuroscience letters, 440*(2), 119-124.

Plailly, J., d'Amato, T., Saoud, M., & Royet, J.-P. (2006). Left temporo-limbic and orbital dysfunction in schizophrenia during odor familiarity and hedonicity judgments. *Neuroimage, 29*(1), 302-313.

Semkovska, M., Bédard, M.-A., Godbout, L., Limoge, F., & Stip, E. (2004). Assessment of executive dysfunction during activities of daily living in schizophrenia. *Schizophrenia research, 69*(2), 289-300.

Simon, J. J., Biller, A., Walther, S., Roesch-Ely, D., Stippich, C., Weisbrod, M., & Kaiser, S. (2010). Neural correlates of reward processing in schizophrenia—relationship to apathy and depression. *Schizophrenia research, 118*(1), 154-161.

Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron, 80*(4), 914-919.

Snaith, R., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *The British Journal of Psychiatry, 167*(1), 99-103.

Strauss, G. P., Frank, M. J., Waltz, J. A., Kasanova, Z., Herbener, E. S., & Gold, J. M. (2011). Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. *Biological psychiatry, 69*(5), 424-431.

Thorndike, E. L. (1933). A proof of the law of effect. *Science*.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review, 55*(4), 189.

Voon, V., Derbyshire, K., Rück, C., Irvine, M., Worbe, Y., Enander, J., . . . Sahakian, B. (2014). Disorders of compulsivity: a common bias towards learning habits. *Molecular psychiatry*.

Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological psychiatry, 62*(7), 756-764.

Waltz, J. A., Schweitzer, J. B., Ross, T. J., Kurup, P. K., Salmeron, B. J., Rose, E. J., . . . Stein, E. A. (2010). Abnormal responses to monetary outcomes in cortex, but not in the basal

ganglia, in schizophrenia. *Neuropsychopharmacology, 35*(12), 2427-2439.

Weickert, T. W., Terrazas, A., Bigelow, L. B., Malley, J. D., Hyde, T., Egan, M. F., . . . Goldberg, T. E. (2002). Habit and skill learning in schizophrenia: evidence of normal striatal processing with abnormal cortical input. *Learning & Memory, 9*(6), 430-442.

Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron, 75*(3), 418-424.

Tables

Table 1: Participant Characteristics

Characteristics	Healthy Controls (N=30)		Individuals with Schizophrenia (N=33)		
Demographics	Mean	SD	Mean	SD	p-value
Age (years)	35.9	8.2	36.7	9.25	0.72
Sex (% male)	46.7%		51.5%		0.70
Ethnicity (% non-Cauasian)	58.0%		56.7%		0.63
Personal Education (years)	15.2	2.7	12.8	2.3	<0.001
Parental Education (years)	14.4	2.6	14.4	3.50	0.97
Medication status					
Atypical antipsychotics (%)	NA		55%		
Typical antipsychotics (%)	NA		3%		
Typical and atypical (%)	NA		6%		
Medicated (no antipsychotics)			15%		
Not Medicated (%)	NA		21%		
Clinical ratings					
Brief Negative Symptom Scale					
Total Score	NA		24.5	13.2	
Avolition/Anhedonia Subscale	NA		16.0	8.8	
Self-Report					
Social anhedonia	8.8	6.0	16.5	6.1	<0.001
Physical anhedonia	11.7	8.5	20.2	8.7	<0.001
Snaith Hamilton Pleasure	50.7	5.9	43.5	10.5	0.001
Neurocognitive Measures					
Wechsler Test of Adult Reading	101.2	14.3	98.3	12.0	0.39

Table 2: Coefficients predicting response repetition from the outcome of the previous trial, the transition type, and diagnostic group.

Coefficient	**Estimate (SE)**	**p-value**
Intercept	1.01 (0.11)	<0.001
Reward	0.99 (0.14)	<0.001
Rarity	-0.20 (0.05)	<0.001
Group	0.01 (0.12)	0.963
Reward x Rarity	0.52 (0.12)	<0.001
Rarity x Group	0.09 (0.05)	0.035
Reward x Group	0.18 (0.13)	0.177
Reward x Rarity x Group	-0.22 (0.11)	0.049

Table 3: External Correlates of model-based and model-free learning

Brief Negative Symptom Scale		
Coefficient	**Estimate (SE)**	**p-value**
Intercept	1.02 (0.15)	<0.001
Reward	1.22 (0.23)	<0.001
Rarity	-0.10 (0.06)	0.09
BNSS	-0.02 (0.01)	0.09
Reward x Rarity	0.25 (0.15)	0.09
Reward x Group	-0.01 (0.02)	0.40
Rarity x Group	0.0003 (0.004)	0.95
Reward x Rarity x Group	0.013 (0.01)	0.24
Intellectual Functioning		
Coefficient	**Estimate (SE)**	**p-value**
Intercept	1.01 (0.12)	<0.001
Reward	1.00 (0.14)	<0.001
Rarity	-0.19 (0.05)	<0.001
Group	0.12 (0.11)	0.28
FSIQ	0.01 (0.009)	0.11
Reward x Rarity	0.50 (0.11)	<0.001
Reward x FSIQ	0.007 (0.01)	0.49
Rarity x FSIQ	-0.01 (0.003)	0.003
Reward x Rarity x FSIQ	0.027 (0.008)	<0.001
Snaith-Hamilton Pleasure Scale (Hedonic Capacity)		
Coefficient	**Estimate (SE)**	**p-value**
Intercept	1.01 (0.12)	<0.001
Reward	1.00 (0.13)	<0.001
Rarity	-0.20 (0.05)	<0.001
Group	0.12 (0.11)	0.28
SNAITH	0.01 (0.01)	0.55
Reward x Rarity	0.51 (0.12)	<0.001
Reward x SNAITH	0.03 (0.02)	0.03
Rarity x SNAITH	-0.01 (0.01)	0.16
Reward x Rarity x SNAITH	0.02 (0.01)	0.11

Table 4: Second stage choice reaction time by group and transition type.

Reaction Time (Stage-2)	Healthy Controls		Schizophrenia Patients	
Transition Type	Mean	SD	Mean	SD
Common Transition (seconds)	0.64	0.17	0.71	0.23
Rare Transition (seconds)	0.77	0.19	0.80	0.26
Rare – Common (seconds)	0.13	0.14	0.9	0.13

Tables

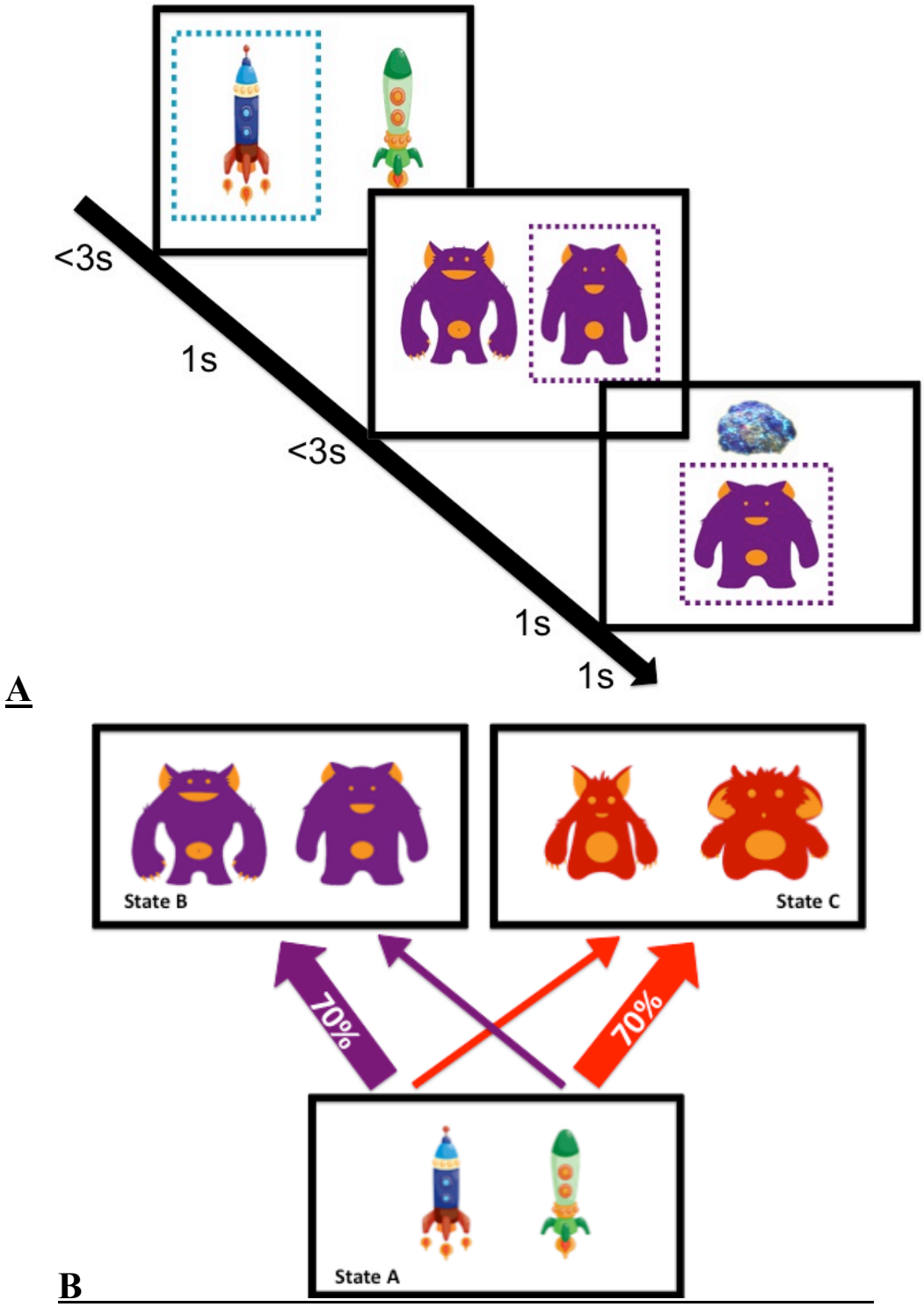

Figure 1: A. Sample trial diagram: binary choice at stage one (spaceships) leads to a second binary choice (aliens) at stage two which is rewarded/non-rewarded (treasure). **B.** Task structure: Each stage one choice predominately leads to one of two second-stage states.

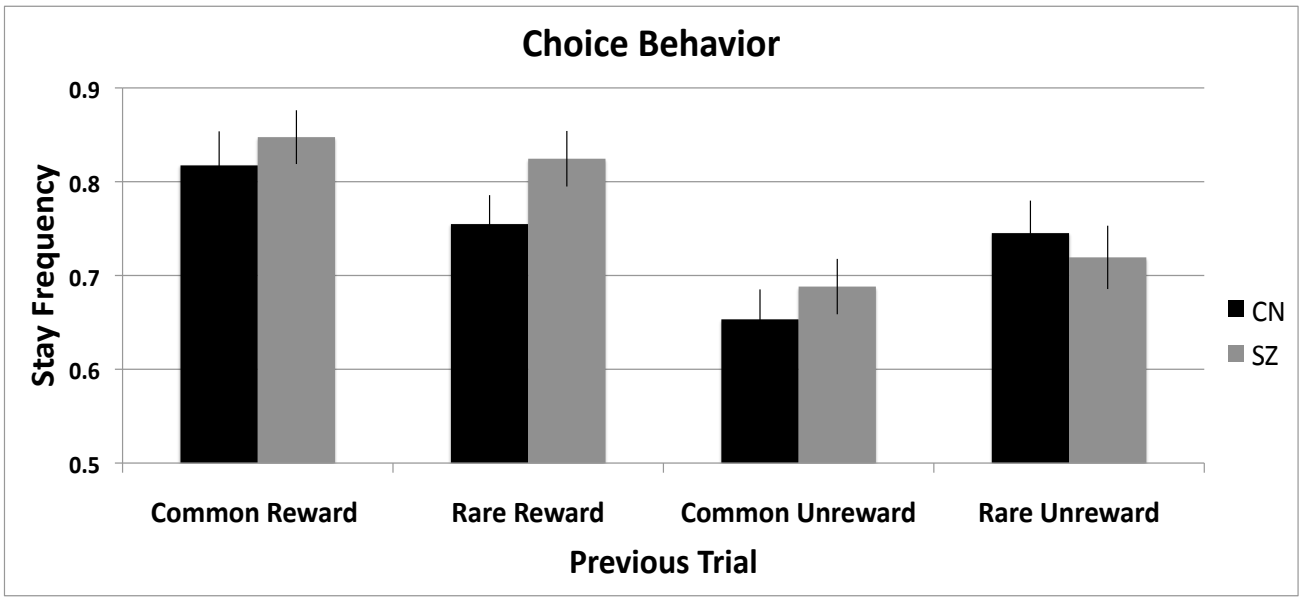

Figure 2: First-stage choice behavior (coded as stay/shift) averaged across individuals within each group.