A computational neuroimaging study of reinforcement learning and goal-directed exploration in schizophrenia spectrum disorders


Abstract

Background. Prior evidence indicates that negative symptom severity and cognitive deficits, in people with schizophrenia (PSZ), relate to measures of reward-seeking and loss-avoidance behavior (implicating the ventral striatum/VS), as well as uncertainty-driven exploration (reliant on rostrolateral prefrontal cortex/rlPFC). While neural correlates of reward-seeking and loss-avoidance have been examined in PSZ, neural correlates of uncertainty-driven exploration have not. Understanding neural correlates of uncertainty-driven exploration is an important next step that could reveal insights to how this mechanism of cognitive and negative symptoms manifest at a neural level.

Methods. We acquired fMRI data from 29 PSZ and 36 controls performing the Temporal Utility Integration decision-making task. Computational analyses estimated parameters corresponding to learning rates for both positive and negative reward prediction errors (RPEs) and the degree to which participants relied on representations of relative uncertainty. Trial-wise estimates of expected value, certainty, and RPEs were generated to model fMRI data.

Results. Behaviorally, PSZ demonstrated reduced reward-seeking behavior compared to controls, and negative symptoms were positively correlated with loss-avoidance behavior. This finding of a bias toward loss avoidance learning in PSZ is consistent with previous work. Surprisingly, neither behavioral measures of exploration nor neural correlates of uncertainty in the rlPFC differed significantly between groups. However, we showed that trial-wise estimates of relative uncertainty in the rlPFC distinguished participants who engaged in exploratory behavior from those who did not. rlPFC activation was positively associated with intellectual function.

Conclusions. These results further elucidate the nature of reinforcement learning and decision-making in PSZ and healthy volunteers.

Introduction

Previous literature indicates that cognitive and motivational systems are impaired in people with schizophrenia (PSZ), such that PSZ exhibit a reduced tendency to engage in goal-directed behavior (Gold, Waltz, Prentice, Morris, & Heerey, 2008; Kring & Barch, 2014). In this literature, goal-directed behavior is frequently examined with paradigms designed to elicit learning from prediction errors (PE), mismatches between expectations and outcomes (McClure, Berns, & Montague, 2003; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006). However, the precise mechanisms that give rise to reduced goal-directed behavior in PSZ remain unknown. Reduced goal-directed behavior could arise from multiple mechanisms including (1) deficient learning from positive PEs (Go-learning); (2) intact or enhanced learning from negative PEs (NoGo-learning); and/or (3) a reduced ability to engage in information-seeking behavior to determine the best actions in unfamiliar environments (diminished uncertainty-driven exploration). Our goal was to examine these contributors to goal-directed behavior and their neural correlates in PSZ and controls.

Goal-directed behavior is commonly examined through paradigms designed to elicit learning for PEs. These tasks are able to quantify the extent to which individuals use surprising positive and/or negative PEs to guide decision-making (Go-learning v. NoGo-Learning). However, PE-driven learning is not the only signal relevant to goal-directed behavior.
For example, in unfamiliar and changing environments individuals must decide between either repeating actions that have resulted in optimal outcomes (exploiting) or trying new actions that could yield even better results (exploring). Uncertainty-driven exploration refers to behaviors where the goal is to increase knowledge of reward contingencies of options about which the least is known, by selecting those options (Frank, Doll, Oas-Terpstra, & Moreno, 2009; Gershman, 2019; Wilson, Geana, White, Ludvig, & Cohen, 2014). In such circumstances, the relative uncertainty about the value of competing options is believed to be a key factor in the trade-off between the exploitation of known contingencies and information seeking about contingencies about which we know little (Cavanagh, Figueroa, Cohen, & Frank, 2012; Frank et al., 2009; Payzan-Lenestour & Bossaerts, 2011). Relative uncertainty is essentially the difference in certainty between ‘known’ and ‘unknown’ options, and thus it is distinct from the mean or overall uncertainty of all options.

Previous reports have provided evidence for deficient learning from positive outcomes, intact learning from negative outcomes, and reduced uncertainty-driven exploration in PSZ, with associations to negative symptoms. Several studies have found that PSZ show deficits in using positive reward PEs to drive behavior, and that such deficits are most robust in those with severe negative symptoms (Gold et al., 2012; Waltz, Frank, Wiecki, & Gold, 2011; Waltz & Gold, 2007). In contrast, learning from negative reward PEs has often been found to be intact in medicated PSZ (Culbreth, Westbrook, Xu, Barch, & Waltz, 2016b; Dowd, Frank, Collins, Gold, & Barch, 2016). Such findings provide an intriguing account of reduced goal-directed behavior wherein PSZ are capable of learning what not to do, to avoid punishment but not what to do, to obtain reward. Finally, multiple studies have found evidence for reduced uncertainty-driven exploration in PSZ (Strauss et al., 2011; Waltz, Wilson, Albrecht, Frank, & Gold, 2020). For example, Strauss found that PSZ exhibited a reduction in goal-directed exploration, which was related to anhedonia (2011). Thus, negative symptoms have been associated with behavioral and neural signals underlying goal-directed behavior in PSZ.

Cognitive deficits in PSZ are also associated with behavioral and neural signals underlying goal-directed behavior. PE-driven learning and uncertainty-driven exploration rely on cognitive resources, in terms of being able to use active representations of value to detect differences among options and changes in environmental contingencies. A consistent finding in the literature is that PSZ show a reduced ability to use cognitively demanding learning strategies compared to controls (Culbreth, Westbrook, Daw, Botvinick, & Barch, 2016a; Gold et al., 2012). Further, reduced uncertainty-driven exploration in PSZ has been linked to cognitive deficits (Waltz et al., 2020).

In addition to detailing potential contributors to goal-directed behavior, work in the basic and clinical sciences has begun to delineate the neural correlates of these behavioral mechanisms. Numerous neuroimaging studies in non-psychiatric samples have shown that intact positive PE-driven learning depends on the intact functioning of ventral striatum (VS) and ventromedial prefrontal cortex (PFC), as these structures are involved in the representation of value and the signaling of positive PEs (Clark, Cools, & Robbins, 2004). Studies examining VS PE-related activation in PSZ have been mixed, some showing impairments in mostly medication-naïve samples (Radua et al., 2015) and others showing intact signaling in medicated patients (Culbreth et al., 2016b). Finally, additional studies in the basic science literature point to a role for rostrolateral (rl) PFC in the representation of relative uncertainty of option value (Badre, Doll, Long, & Frank, 2012; Zajkowski, Kossut, & Wilson, 2017). Importantly, the neural correlates of exploratory behavior have yet to be examined in PSZ.

Our goal was to examine contributors to goal-directed behavior and their neural correlates in PSZ and controls. Behaviorally, we hypothesized that impairments in the execution of goal-directed behavior observed in psychosis will be characterized by both (1) impaired learning from positive PEs and (2) intact learning from negative PEs. However, we hypothesized that impairments in goal-directed behavior would extend beyond reduced PE-driven learning, involving mechanisms for adjudicating the explore/exploit trade-off. We employed a behavioral paradigm, the Temporal Utility Integration (TUI) task, which lends itself to computational modeling, allowing us to estimate parameters corresponding to learning from positive and negative outcomes, as well as exploratory behavior. Computational modeling also enabled us to estimate the contributions of representations of certainty about value in modulating decision-making.

Administration of our experimental task in conjunction with fMRI allowed for examination of the neural correlates associated with aspects of goal-directed behavior. Importantly, this is the first study to investigate neural mechanisms of exploratory behavior in PSZ. We predicted that activity in rlPFC would track the tendency to use relative uncertainty to drive exploration (distinguishing individuals prone to use exploration from those who do not). Finally, we expected neural correlates to systematically relate to the severity of motivational deficits and cognitive impairment in PSZ.

Methods and materials

Recruiting and screening

Twenty-nine PSZ and 36 demographically matched controls provided written informed consent to protocols approved by the Institutional Review Board of the University of Maryland School of Medicine (Protocol HP-00051996) and successfully completed a behavioral task in the MRI scanner. All patients were chronic outpatients on stable antipsychotic medication regimens (no changes for four weeks). Diagnosis of schizophrenia or schizoaffective disorder in patients was confirmed using the Structured Clinical Interview for DSM-IV-R Disorders (First & Gibbon, 2004), as was the absence of Axis I and Axis II diagnoses in controls. Additional exclusionary criteria included pregnancy and admission of past substance dependence.

Assessment

Premorbid intellectual function was assessed using the Wechsler Test of Adult Reading (WTAR; Weschler, 2001). Standard symptom ratings were obtained for PSZ using the Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1989), the Brief Psychiatric Rating Scale (Overall & Gorham, 1962), the Calgary Depression Scale (Addington, Addington, Maticka-Tyndale, & Joyce, 1992), and the Young Mania Rating Scale (Young, Biggs, Zigler, & Meyer, 1978). We computed the experiential negative symptom factor score by averaging item scores for avolition/role-functioning and anhedonia/asoiality from the SANS (Table 1). We computed individual psychosis scores for PSZ by averaging scores from the four psychosis items
Table 1. Demographic, clinical, and standard cognitive characterization of participants

<table>
<thead>
<tr>
<th></th>
<th>Schizophrenia Patients (N = 29)</th>
<th>Controls (N = 36)</th>
<th>Sig. of group diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>38.0 (11.7)</td>
<td>37.3 (11.7)</td>
<td>p = 0.797</td>
</tr>
<tr>
<td>Gender</td>
<td>34.5% F</td>
<td>30.6% F</td>
<td>p = 0.794</td>
</tr>
<tr>
<td>Race</td>
<td>31.0% NW</td>
<td>38.9% NW</td>
<td>p = 0.605</td>
</tr>
<tr>
<td>Nicotine use</td>
<td>27.6% Y</td>
<td>11.8% Y</td>
<td>p = 0.197</td>
</tr>
<tr>
<td>Subject education (years)</td>
<td>13.3 (1.9)</td>
<td>15.0 (1.9)</td>
<td>p = 0.001</td>
</tr>
<tr>
<td>Parental education (years)</td>
<td>14.1 (2.6)</td>
<td>14.7 (2.6)</td>
<td>p = 0.417</td>
</tr>
<tr>
<td>Neuropsychological testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTAR scaled score</td>
<td>105.0 (19.7)</td>
<td>113.9 (9.7)</td>
<td>p = 0.032</td>
</tr>
<tr>
<td>Symptom ratings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BPRS item score</td>
<td>1.53 (0.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BPRS psychosis item score</td>
<td>1.64 (0.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean SANS item score</td>
<td>1.49 (0.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean SANS Avo/Anhed item score</td>
<td>1.90 (1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean CDS total score</td>
<td>1.31 (1.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean YMRS total score</td>
<td>1.62 (3.06)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WASI, Wechsler Abbreviated Scale of Intelligence; WTAR, Wechsler Test of Adult Reading; WRAT4, Wide-ranging Achievement Test, Reading Subtest; BPRS, Brief Psychiatric Rating Scale; SANS, Scale for the Assessment of Negative Symptoms; Avo/Anhed, Avolition/Anhedonia subscales; CDS, Calgary Depression Scale; YMRS, Young Mania Rating Scale.

from the BPRS (Suspiciousness, Grandiosity, Hallucinations, and Unusual Thought Content).

**Experimental task**

In the TUI task (Frank et al., 2009), participants observe a clock hand which completes a single revolution over 5 s (Fig. 1). Participants were instructed to stop the clock hand by pressing a response button, after which a number of points were awarded (see online Supplemental Methods for verbatim instructions). The outcomes of trials (i.e. points earned) were displayed on the center of the clock face at the end of each trial. Participants could only win points on trials on which they responded in less than 5 s.

Points were awarded with a probability and magnitude that varied as a function of response time (RT; Fig. 1). The Matlab code used to generate reward probability and magnitude, across time, within a trial, is provided in online Supplemental Methods S2. In three conditions, reward probability decreased with RT and reward magnitude increased with RT. However, the precise relationships between reward probability and RT, and between reward magnitude and RT, varied across conditions such that the product of reward probability and magnitude, expected value (EV), increased with increasing RT in one condition (the increasing expected value/IEV condition), decreased in another (the decreasing expected value/DEV condition), and remained constant in a third (the constant expected value/CEV condition). In a fourth condition (the constant expected value – reversed/CEVr condition), EV remained constant, but reward probability increased with RT and reward magnitude decreased with RT. The CEV and CEVr conditions served as control conditions. Because expected value is higher earlier in the clock, in the DEV condition, subjects are more likely to receive a positive PE for speeding up responding in the DEV condition; thus, the DEV condition primarily assesses the degree to which people learn to speed up responding, following experience of positive PEs. Because expected value is higher later in the clock, in the IEV condition, subjects are more likely to receive a negative PE for speeding up responding in the IEV condition; thus, the IEV condition primarily assesses the degree to which people learn to slow down responding, following experience of negative PEs. Participants performed two blocks each of the DEV and IEV conditions, and one block each of the CEV and CEVr conditions, for a total of six 40-trial blocks. The orders of these blocks were randomized across participants.

**Behavioral data analysis**

Behavioral measures of performance are shown in Table 2. For each subject, we computed mean RT changes in each condition (first 10 trials–last 10 trials of each block). We also averaged RT differences between the previous trial and the current trial, as a function of the previous trial’s outcome (wins: >0 points; non-wins: 0 points). This was done separately in the IEV and DEV conditions and was used as a behavioral measure of exploration. We refer to this variable as ‘RT swing’. Finally, we compared groups on the total number of points earned (online Supplementary Table S1). Groups did not differ in number of points earned for either the IEV or DEV conditions.

**Computational modeling**

**Original model**

We used a previously validated computational model of the TUI paradigm to probe contributors of goal-directed behavior on a subject-wise basis (Badre et al., 2012; Frank et al., 2009; Strauss et al., 2011). This model estimates the degree to which individuals modulate RT as a function of positive and negative PEs, as well as uncertainty-driven exploration.

On each trial (t), our model estimates a response time, \( \hat{RT}(t) \):

\[
\hat{RT}(t) = K + \lambda RT(t - 1) + \nu[(RT_{best} - RT_{avg})] - Go(t) + NoGo(t) + \rho[\mu_{slow}(t) - \mu_{fast}(t)] + \epsilon[\theta_{slow}(t) - \theta_{fast}(t)]
\]

Here, \( K \) is a free parameter capturing a participant’s baseline response speed, \( \lambda \) represents the autocorrelation between the previous and current RT, and \( \nu \) captures the tendency of individuals to adapt RT toward the single largest reward experienced thus far, \( RT_{best} \), the ‘Going for the Gold’ parameter (Badre et al., 2012; Frank et al., 2009).

To probe contributions of exploitative and exploratory behavior, the model includes components that track the expected value (V) of two specific classes of actions separately, fast and slow RT, compared to a participant’s local average RT. The local average RT (\( RT_{avg} \)) is calculated as follows:

\[
RT_{avg}(t) = RT_{avg}(t - 1) + \alpha[RT(t - 1) - RT_{avg}(t - 1)]
\]
Even though RT is continuous, the reward functions are monotonic. Subjects are told that sometimes it will be better to respond faster and sometimes slower. Thus, participants only need to track the reward value of responding relatively faster for slower and adjust RT accordingly.

For both exploitative and exploratory components, value is updated using a delta rule. \( \alpha \) is the rate at which new information is integrated into \( V \) and \( \delta \) is the reward PE [\( \text{Outcome}(t-1) - V(t-1) \)].

\[
V(t) = V(t-1) + \alpha \delta(t-1)
\]

The exploitative component of the model tracks the expected value associated with both fast and slow RT, allowing participants to continuously modulate RT in proportion to their relative difference. More specifically, the model assumes that participants track the probability of obtaining a better than average outcome following fast or slow responses, which are separately computed via Bayesian integration:

\[
P(\theta|\delta_1 \ldots \delta_n) \propto P(\delta_1 \ldots \delta_n|\theta)P(\theta)
\]

Here, \( \theta \) represents the parameters in the probability distribution and \( \delta_1 \ldots \delta_n \) represents the prediction errors observed thus far in the experiment. The difference in expected value means (\( \mu_{\text{slow}} \), \( \mu_{\text{fast}} \)) contributes to RT on trial \( t \) scaled by free parameter \( \rho \):

\[
\rho[\mu_{\text{slow}}(t) - \mu_{\text{fast}}(t)]
\]

The exploratory component of the model capitalizes on the uncertainty of the probability distributions to strategically explore those responses for which reward statistics are most uncertain. Specifically, the model assumes that subjects explore uncertain RTs to reduce uncertainty. This component is computed as:

\[
\epsilon[\sigma_{\text{slow}}(t) - \sigma_{\text{fast}}(t)]
\]

Fig. 1. Temporal Utility Integration (TUI) task. (a) Example clock-face stimulus; (b) the probability of reward occurring as a function of response time; (c) reward magnitude (contingent on RT); (d) expected value across trials for each time point. The functions are designed such that the expected value in the beginning in DEV is approximately equal to that at the end in IEV so that if optimal, subjects should obtain the same average reward in both IEV and DEV. Within a given condition (consisting of a block of 40 trials) individuals can learn to produce responses that yield the most points, on average, by stopping the clock at the optimal time (with a key press). Without exploring multiple RTs, a subject might be reinforced often for a given RT, but never discover whether he/she might obtain a larger number of points if he/she had only explored other options. Note that CEV and CVER have the same EV, so the black line represents EV for both conditions. The x-axis in all plots corresponds to the time after onset of the clock stimulus at which the response is made. Reprinted from Ref. 1.
Here $\sigma_{\text{slow}}$ and $\sigma_{\text{fast}}$ are uncertainties, quantified in terms of standard deviations of the probability distributions tracked by the Bayesian update rule, and $\epsilon$ is a free parameter controlling the degree to which subjects make exploratory responses in proportion to relative uncertainty, $\sigma_{\text{slow}}(t) - \sigma_{\text{fast}}(t)$.

Finally, Go and NoGo learning reflect a striatal bias to speed responding as a function of positive RPEs and to slow responding as a function of negative RPEs. Evidence for speeding and slowing in the task is separately tracked:

$$\text{Go}(t) = \text{Go}(t-1) + \alpha_{G}\delta_+(t-1)$$
$$\text{NoGo}(t) = \text{NoGo}(t-1) + \alpha_{N}\delta_-(t-1)$$

where $\alpha_{G}$ and $\alpha_{N}$ are learning rates scaling the effects of positive ($\delta_+$) and negative ($\delta_-$) errors in expected value prediction $V$ (i.e. positive and negative RPE).

**Sticky choice model**

We compared the model described above with a model testing the alternative idea that people are averse to uncertainty. To test uncertainty aversion, $\epsilon$ was allowed to take on a negative value. To aid interpretation of uncertainty aversion $\nu$, simply mimicking the tendency to respond the same as previous trials, we added an additional free parameter, **sticky choice**, where $\lambda RT(t-1)$ is replaced with $\lambda$ *sticky*(t). This allowed for estimation of a decaying function of previous trials’ RTs $v.$ simply accounting for the previous trials’ RTs:

$$\text{sticky}(t) = \text{RT}(t-1) + d \times \text{sticky}(t-1)$$

Here, $d$ is a decay parameter influencing the degree to which prior RTs influence the current RT estimate. To limit the number of free parameters in the sticky choice model, we removed the ‘going for the gold’ parameter, as it was not critical for hypotheses examining effort aversion. See online Supplementary Table S2 for model comparison. Both models appeared to fit participant behavior in roughly an equivalent manner. In the following analyses, we report results from the original model; however, results from the sticky choice model can be found in the online Supplemental Data S1.

**Modeling summary**

The three parameters of interest in our study are $\alpha_{P}$, $\alpha_{N}$, and $\epsilon$ (Table 2). Similar to Badre et al. (2012), the estimation of the $\epsilon$ parameter also allowed us to classify participants as ‘Explorers’ and ‘Non-explorers’, based on whether this $\epsilon$ parameter was positive or not. Along with fitting parameters estimated for each participant, the computational model also provides trial-level estimates of expected value, relative uncertainty (the difference in uncertainty between fast v. slow RT), and prediction error. We use these trial-level variables in the neuroimaging analyses described below to elicit the neural correlates of decision variables. See online Supplementary Fig. S1 for correlations among trial-level variables.

**Analyses of event-related MRI data**

Based on prior literature (Badre et al., 2012), we examined effects of relative uncertainty in right rLPFC. Taken from Badre et al., the Talairach coordinates were $[27 50 23]$ (2012). The ROI was a sphere of 10 mm radius. Finally, we specified an ROI in the VS, to investigate RPE-evoked activity, consisting of two spheres of 5 mm radius centered on Talairach coordinates: $[\pm10 8 -4]$ (Culbreth et al., 2016b; Schlagenhauf et al., 2014).

In each ROI, we extracted mean beta values for the two regressors of interest described above (RPE and relative uncertainty) and then performed one-sample $t$ tests comparing the overall mean contrast to zero, as well as two-sample $t$ tests to test for between-group differences. In order to examine how symptom...
severity and measures of intellectual function modulated MRI responses, we performed Spearman correlations on the mean beta values from the ROIs. To characterize the influence of antipsychotic drugs on BOLD signal responses in ROIs, we converted antipsychotic doses for PSZ to haloperidol equivalents (Andreasen, Pressler, Nopoulos, Miller, & Ho, 2010).

As stated above, our computational modeling analyses allowed us to characterize participants as ‘Explorers’ v. ‘Non-Explorers’. In order to assess effects of diagnosis, premorbid IQ, and regional brain activity on explorer status, we performed logistic regression analyses with predictor variables including diagnosis, premorbid IQ, and relative-uncertainty-evoked rIPFC activity and dependent variable of explorer status.

**Results**

**Behavioral measures of goal-directed behavior**

**PE-driven learning**

The DEV condition assesses Go-Learning, while the IEV condition assesses NoGo-Learning (Frank et al., 2009; Moustafa, Cohen, Sherman, & Frank, 2008). Analyses revealed that PSZ showed reduced DEV acceleration, relative to controls ($F_{3,186}$ for group × block interaction = 2.981, $p = 0.033$), whereas groups did not differ in RT deceleration in the IEV condition ($F_{3,186}$ for group × block interaction = 0.437, $p = 0.727$; Fig. 2). Patients and controls also did not differ significantly in the control conditions (CEV and CEVr). Patients and controls did not differ in the total number of points earned in DEV conditions, IEV conditions, or across all six trial blocks (online Supplementary Table S1). In summary, the current results provide further evidence for reduced learning from positive outcomes and intact learning from negative outcomes in medicated PSZ. See online Supplementary Fig. S2 for raw trial-wise RT by trial number by condition.

In terms of symptom associations, motivational deficits and premorbid IQ correlated positively with IEV deceleration (Fig. 3; Table 3). Patients with the greatest motivational deficits and highest premorbid IQs were the most sensitive to loss. We observed no significant correlations between RL measures and positive symptom severity (Table 3).

**Exploratory behavior**

Averaged RT differences between the previous trial and the current trial (‘RT swings’) served as our behavioral measure of exploration. Across groups, participants showed large RT swings following non-win outcomes, suggesting exploration after suboptimal actions. Patients and controls did not differ in their mean no-win RT swings in the IEV and DEV conditions, or in mean win-shifts in the IEV condition (online Supplementary Table S3). However, we observed that controls showed larger mean win-shifts compared to PSZ in the DEV condition ($t_{41} = 2.222, p = 0.030$). When we examined relationships between clinical measures and experimental measures of exploration, we observed that mean RT shifts following non-wins, in the IEV condition, positively correlated with the severity of motivational deficits in PSZ (Table 4). In both PSZ and controls, mean RT shifts following IEV non-wins positively correlated with premorbid IQ.

**Computational measures of goal-directed behavior**

Computational modeling analyses recapitulated findings from behavioral measures of goal-directed behavior (online Supplementary Fig. S3). Specifically, the contrast in learning rates ($[α_0 – α_3]$) correlated positively with the [DEV acceleration–IEV deceleration] contrast in both PSZ ($ρ = 0.580, p = 0.001$) and controls ($ρ = 0.342, p = 0.048$; online Supplementary Table S4). These associations lend credence to the idea that DEV acceleration reflects a bias toward greater learning from positive RPEs and that IEV deceleration reflects a bias toward greater learning from negative RPEs. However, the contrast in learning rates for positive and negative RPEs ($[α_3 – α_0]$) did not differ significantly between groups (online Supplementary Table S5).

Surprisingly, we did not observe significant group differences or negative symptom associations when examining the $ε$ parameter, our computational estimate of uncertainty-driven exploration (Table 4; online Supplementary Table S5). However, $ε$ was positively associated with premorbid IQ, a measure of intellectual function. Further, we conducted analyses where participants were classified as ‘Explorers’ v. ‘Non-Explorers’ based on $ε > 0$. Here, we observed that premorbid IQ showed a significant positive association with explorer status, but diagnosis and their interaction did not (beta = 0.12, $p$ value = 0.04; online Supplementary Table S6). Thus, measures of intellectual function were positively associated with exploration measures, but diagnosis was not.

**Neural correlates of goal-directed behavior**

**PE-driven learning**

Given that prior research has demonstrated a role of the VS in the representation of value and signaling of RPEs, we hypothesized a similar finding in the present study. Indeed, the results of ROI analyses were indicative of the VS positively tracking RPE magnitude and valence ($t_{41} = 4.227, p < 0.01$; online Supplementary Fig. S4). Of note, a stronger VS RPE signal was observed in controls who showed greater DEV acceleration (but not PSZ; online Supplementary Table S7). However, there was no significant effect of diagnosis on RPE-related VS activity ($t_{41} = −0.951, p = 0.346$) (Fig. 4).

**Neural correlates of uncertainty-driven exploration**

Based on Badre et al. (2012), we hypothesized that rIPFC activity would track the tendency to use relative uncertainty to drive exploration, thus distinguishing ‘Explorers’ from ‘Non-explorers’. Regression analyses utilizing diagnosis and rIPFC activity to predict explorer status yielded an overall model at a trend level of significance ($χ^2 = 7.14, p = 0.07$). The model indicated that rIPFC activity and the interaction between diagnosis and rIPFC were significant predictors ($p = 0.04$ and $p = 0.03$, respectively). Diagnosis alone, however, was not significant ($p = 0.68$).

Results from a regression model that included diagnosis and premorbid IQ estimates (from the WTRT), as predictors, as well as the interaction term, was significant ($F_{3,57} = 2.95, p = 0.040$) and indicated that both diagnosis and premorbid IQ were significant positive predictors of rIPFC activity (online Supplementary Table S8). Their interaction was observed at a trend level of significance. These analyses indicate that both diagnosis and cognitive measures are associated with rIPFC relative uncertainty to drive exploration activity. No relationship was observed between measures of motivation deficits and rIPFC activity. We observed no significant correlations between ratings
of positive symptom severity from the BPRS and behavioral, modeling, or neural measures of exploration.

**Medication effects**

Spearman correlation analyses revealed that antipsychotic medication dose was not significantly related to any measures of interest pertaining to symptoms, behavior, or neural activity among PSZ (online Supplementary Table S9).

**Additional models and fit analyses**

While we present analyses of the original model in the main manuscript, analyses in a subset of participants who showed good fit to the computational model, as well as analyses using the 'sticky choice' model can be found in the online Supplemental Results (Table S14). Results were largely similar.

**Discussion**

We examined contributors to goal-directed behavior and their neural correlates in PSZ and controls. Specifically, we probed (1) learning from positive PEs; (2) learning from negative PEs; and (3) uncertainty-driven exploration. Importantly, this was the first study to examine neural correlates of uncertainty-driven exploration in PSZ. Consistent with previous reports, PSZ behaviorally demonstrated reduced reward-seeking behavior (decreased DEV acceleration) (Gold et al., 2012; Strauss et al., 2011). In contrast, negative symptoms were positively correlated with IEV deceleration, such that patients with more severe motivational deficits showed greater IEV deceleration (enhanced NoGo-learning). This finding of a stronger bias toward NoGo-learning, in PSZ, relative to controls, is also consistent with our previous work (Gold et al., 2012; Waltz et al., 2011; Waltz & Gold, 2007). Surprisingly, behavioral measures of uncertainty-driven exploration did not differ significantly.

---

**Fig. 2.** Behavioral performance of patients and controls in the four experimental conditions. (a) Relative to controls, PSZ showed reduced DEV acceleration, from the first to the last block of trials, whereas patients and controls did not differ in their increases in mean response latencies, from the first to the last block of trials, in the (b) IEV, (c) CEV, or (d) CEVr conditions.
between groups; however, exploration was positively associated with intellectual function. At a neural level, we found clear evidence of RPE signals in VS, but no between-group differences. Replicating Badre et al. (2012), we showed that trial-wise estimates of relative uncertainty in the rLPFC distinguished participants who engaged in exploratory behavior from those that did not. This finding is consistent with work from Zajkowski et al., showing that rLPFC intervention impairs directed exploration (Zajkowski et al., 2017). Further, mirroring behavioral analyses, uncertainty-related activation in rLPFC was positively associated with intellectual function. Unexpectedly, however, neural correlates of uncertainty in the rLPFC did not differ between groups.

Regarding exploration, we replicated prior modeling results indicating that representations of uncertainty contribute to decisions to sample options about which less is known, and that rLPFC plays a role in this process. While we observed no between-group difference in the exploration parameter, $\epsilon$, and no significant correlations between clinical ratings for motivational deficits and $\epsilon$ (in contrast to Strauss et al., 2011), we observed effects of cognition on both $\epsilon$ and neural activity associated with uncertainty-driven exploration. Our finding of relationships between measures of intellectual function and goal-directed exploration indicates that those with decreased cognitive capacity are less likely to engage in uncertainty-driven exploration and is consistent with recent results from our group investigating uncertainty-driven exploration using a different paradigm (Waltz et al., 2020).

There are several explanations for null findings, regarding effects of diagnosis and negative symptom severity on uncertainty-driven exploration. First, reduced uncertainty-driven exploration may be characteristic of only a subset of PSZ. Thus,
the effect may depend on the particular sample recruited. In our sample, participants exhibited relatively high cognition and relatively low negative symptoms. Second, the TUI task is not designed to isolate directed and random components of exploration. Whereas directed exploration is driven by information seeking, ‘random exploration’ pertains to behavioral variability that drives exploration by chance (Wilson et al., 2014). In our paper using a different paradigm (Waltz et al., 2020), we found that PSZ and controls differed on measures of directed exploration, but not measures of random exploration. Third, some PSZ may generate relatively accurate representations of uncertainty, which then do not influence decision-making or even lead to the active avoidance of more uncertain options—a phenomenon called ambiguity aversion. Extreme ambiguity aversion was characteristic of a substantial fraction of PSZ in our recent behavioral study (Waltz et al., 2020). It is possible that cognitive functioning may have a stronger association with uncertainty-driven exploration and reinforcement learning, compared to negative symptomatology, than we previously appreciated.

Consistent with Badre et al. (2012), neural activity in rPFC distinguished ‘Explorers’ from ‘Non-explorers’ in controls but not PSZ. This may result from the fact that uncertainty, even when adaptively represented, did not always lead to exploratory behavior in patients (sometimes, it led to the active avoidance of more uncertain options). Thus, uncertainty representations in
the brain may have been less coupled from exploration-related neural activity in rLPFC. However, due to the fact that uncertainty-driven exploration was most characteristic of participants with higher levels of cognitive functions, it is understandable that greater relative-uncertainty-driven activity in rLPFC was observed in those with higher intellectual functioning, overall.

Limitations
A modest sample size may have precluded the detection of behavioral and neural associations of goal-directed exploration in the present sample. Additionally, PSZ recruited for this study consisted of stable outpatients on antipsychotic medication. We also did not assess duration of illness. Although we observed no effects of standardized antipsychotic dose, possible effects of medication cannot be ruled out.

Summary
Schizophrenia patients demonstrated reduced reward-seeking behavior and showed IEV deceleration that correlated positively with experiential negative symptoms. Surprisingly, behavioral measures of uncertainty-driven exploration were not significantly different between groups. We showed that trial-wise estimates of relative uncertainty in the rLPFC distinguished Explorers and Non-Explorers. Uncertainty-related activation in rLPFC was also positively associated with intellectual function. These results further elucidate the nature of reinforcement learning and decision-making in PSZ and controls, linking specific cognitive and computational processes to specific neural substrates, which could serve as biomarkers to quantify the effects of potential interventions.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291722003993

Conflict of interest. The authors report no conflicts of interest related to the current manuscript.

References


