Deficits in Positive Reinforcement Learning and Uncertainty-Driven Exploration Are Associated with Distinct Aspects of Negative Symptoms in Schizophrenia

Gregory P. Strauss, Michael J. Frank, James A. Waltz, Zuzana Kasanova, Ellen S. Herbener, and James M. Gold

Background: Negative symptoms are core features of schizophrenia (SZ); however, the cognitive and neural basis for individual negative symptom domains remains unclear. Converging evidence suggests a role for striatal and prefrontal dopamine in reward learning and the exploration of actions that might produce outcomes that are better than the status quo. The current study examines whether deficits in reinforcement learning and uncertainty-driven exploration predict specific negative symptom domains.

Methods: We administered a temporal decision-making task, which required trial-by-trial adjustment of reaction time to maximize reward receipt, to 51 patients with SZ and 39 age-matched healthy control subjects. Task conditions were designed such that expected value (probability × magnitude) increased, decreased, or remained constant with increasing response times. Computational analyses were applied to estimate the degree to which trial-by-trial responses are influenced by reinforcement history.

Results: Individuals with SZ showed impaired Go learning but intact NoGo learning relative to control subjects. These effects were most pronounced in patients with higher levels of negative symptoms. Uncertainty-based exploration was substantially reduced in individuals with SZ and selectively correlated with clinical ratings of anhedonia.

Conclusions: Schizophrenia patients, particularly those with high negative symptoms, failed to speed reaction times to increase positive outcomes and showed reduced tendency to explore when alternative actions could lead to better outcomes than the status quo. Results are interpreted in the context of current computational, genetic, and pharmacological data supporting the roles of striatal and prefrontal dopamine in these processes.

Key Words: Computational model, dopamine, negative symptoms, reinforcement learning, reward, schizophrenia

Dopaminergic (DA) signaling plays a key role in the detection, evaluation, and prediction of rewards. Several structures that receive DA input are differentially involved in specific aspects of reward learning. For example, the striatum and orbitofrontal cortex have been found to be involved in reward prediction and reward-based decision-making, with the orbitofrontal cortex being particularly responsive to reward magnitudes (1–6). In reinforcement learning models of corticostriatal circuitry (7,8), phasic DA signals are proposed to modify synaptic plasticity in the corticostriatal pathway (9,10) and subsequently reinforce “Go” (learning to pursue actions that have high reward probability) and “NoGo” learning (learning to avoid actions with low reward probabilities) (7,11). Specifically, increases in phasic striatal DA support Go learning from positive feedback via D1 receptor stimulation, whereas decreases in phasic striatal DA support avoidance learning from negative feedback via D2 receptor disinhibition.

This model has been applied to understand patterns of reward learning in Parkinson’s disease patients (7), who have depleted striatal DA levels because of the disease but increased striatal DA levels after DA medication (12,13). Supporting the models, it has been found that Parkinson’s disease patients taking medication learn better from positive than negative decision outcomes, but patients not taking medication showed the opposite bias (8,14–16). Imaging studies show that these biases are accompanied by medication-induced increased sensitivity to positive prediction errors and reduced sensitivity to negative prediction errors in the ventral and dorsolateral striatum (17). Behaviorally, these medication-induced effects have been primarily observed in tasks where participants learned stimulus-response relationships as a function of reinforcement (e.g., probabilistic learning).

Recent studies have found that schizophrenia (SZ) patients exhibit reinforcement learning abnormalities, specifically in learning to integrate the history of probabilistic positive decisions across trials (18–20). These deficits have been attributed to deficient phasic striatal DA signals and D1 receptor functionality, leading to poor Go learning. A deficit in learning to repeat those actions most likely to yield positive reinforcement might provide an intuitively appealing account that could explain negative symptoms. Several studies show that negative symptoms are associated with impairments in rapid, trial-to-trial behavioral adaptation in response to recent changes in reinforcement values, particularly during early phases of learning. We have argued that this deficit in rapid acquisition associated with negative symptoms is likely to stem from prefrontal cortical dysfunction (19,20). Similar patterns of early learning deficits are seen in patients with orbito-
frontal damage (21) and in healthy participants with the val/val genotype of the catechol-O-methyltransferase (COMT) gene (15), who have reduced prefrontal (and particularly, orbitofrontal) DA levels (22). However, relationships have also been reported between negative symptoms and deficient reward-related blood oxygen level dependent activity in basal ganglia (BG) structures. These changes in BG blood oxygen level dependent signal do not themselves necessarily indicate that the BG are the source of the deficits. For example, reduced top-down input from orbitofrontal cortex to BG during reward would result in reduced BG activations. Thus, the neural underpinnings of reinforcement learning deficits and negative symptoms are at present unclear.

In the current study, we examined reinforcement learning in SZ with a task that required subjects to modulate the response time of a single motor response, where both reinforcement probabilities and magnitudes varied as a function of reaction time (RT) (23). Behavioral task conditions were designed to assess Go learning, NoGo learning, and relative sensitivity to reward frequency versus magnitude.

Computational modeling was performed in an effort to obtain a richer understanding of the behavioral findings. The modeling approach allows us to assess the degree to which participants adjust response times as a function of positive and negative prediction errors across all four of the task conditions (not solely in the conditions where it is most advantageous to do so), while distinguishing these measures from several other components to RT adjustment in the task (captured by other model parameters).

We hypothesized, on the basis of theories suggesting that SZ is associated with abnormal DA signaling (high tonic DA but low phasic DA) and impaired D1 function in particular (25–26), that SZ patients would fail to show relative speeding when rewards were most available for fast responses, and thus earlier responding results in better-than-expected outcomes (i.e., positive prediction errors). Conversely, we hypothesized, on the basis of our previous findings (19) and because patients were medicated with D2 antagonists, that SZ patients would show intact slowing as a function of negative prediction errors when early responses produced outcomes that were lower than expected on average. The latter interpretation is supported by computational simulations showing that D2 blockade enhances NoGo learning and RT slowing (see Wiericki et al. [27]) and recent demonstration that D2 blockade enhances NoGo learning in Tourette’s syndrome patients (16).

Evidence for such a pattern of spared sensitivity to negative outcomes, coupled with reduced ability to learn to approach responses leading to positive reinforcement could be considered a perfect neurobehavioral recipe for avolition and anhedonia. In light of previous studies indicating that reinforcement learning impairments are most severe in high negative symptom patients (28,29), we therefore also examined the role of negative symptoms in Go and NoGo learning, with the prediction that high negative symptom patients would show the greatest Go learning impairment and comparatively intact NoGo performance. Alternatively, rather than resulting from reinforcement learning deficits per se, some aspects of negative symptoms might be characterized by a reduced tendency to appropriately explore alternative actions in the hope that they might produce better outcomes. Notably, the computations of outcome uncertainty used to guide exploration are thought to depend on neuromodulation within the prefrontal cortex (30–32). Accordingly, a recent genetic study with the same task used here showed that, consistent with striatal DA genetic effects on Go/NoGo learning, individual differences in uncertainty-driven exploration were predicted by COMT val/met genotype (32). We thus applied the same computational analyses of trial-by-trial responses in the current study to investigate the prediction that patients would show not only differences in speeding and slowing as a function of prediction errors but also whether they would exhibit uncertainty-driven exploration.

Methods and Materials

Participants
Participants included 51 patients meeting DSM-IV-TR criteria for SZ and 39 healthy control subjects (CN). The patients were recruited from the outpatient clinics at the Maryland Psychiatric Research Center and were studied during a period of clinical stability. All patients met DSM-IV diagnostic criteria for SZ or schizoaffective disorder. Consensus diagnosis was established with a best-estimate approach on the basis of medical records and confirmed with the Structured Clinical Interview for DSM-IV (SCID) (33). All patients were receiving antipsychotic medications.

Control subjects were recruited through random digit dialing and word of mouth among individuals recruited through random digit dialing. All CN underwent a screening interview and denied lifetime and family history of psychosis and any active Axis I disorder on the SCID. All patients denied lifetime history of significant neurological conditions and recent substance abuse as determined by the SCID (none within 6 months). Upon entry to our subject pool, we routinely screen for substance use via urine toxicology testing. In the current study, targeted urine toxicology testing was performed in instances where there were suspicions of substance use. Patient and control groups did not significantly differ in age, parental education, gender, or ethnicity. Patients had fewer years of total education and lower Wechsler Abbreviated Scale of Intelligence estimated full-scale IQs than CN (Table 1).

Schizophrenia patients were also divided into high (HI-NEG) and low negative (LOW-NEG) symptom groups on the basis of a median split on the Scale for the Assessment of Negative Symptoms (SANS) (34,35) total score. The 22-item version of the SANS developed in the CONSIST clinical trial (Cognitive and Negative Symptoms in Schizophrenia Trial) was used (35), which has fewer items than the original 30- or 25-item version, with total scores ranging from 0 to 110. The three groups did not significantly differ on age, parental education, gender, or ethnicity; however, they did differ on IQ, such that CN had significantly higher IQ than both SZ groups. There were no differences in IQ between the HI-NEG and LOW-NEG patients. The HI-NEG and LOW-NEG patients significantly differed on the Brief Psychiatric Rating Scale negative symptom factor score, but not on positive symptoms, disorganization, or total scale score. HI-NEG and LOW-NEG patients were also prescribed a similar regimen of antipsychotic medications at the time of testing and did not differ on chlorpromazine (CPZ) equivalent dosage (36) (Table 1).

General Procedures

The current tests were administered as part of a larger battery of reward-learning, symptom interview, and neuropsychological measures. For each subject, demographic, diagnostic, and symptom ratings were completed before administration of the neuropsychological evaluations. Symptom interviews included the SANS and Brief Psychiatric Rating Scale (37). Patients and control participants recruited from the community received monetary compensation for participation. Study personnel administering the neuropsychological tasks included BA- and MA-level research assistants.

Temporal Utility Integration Task

Participants completed the “temporal utility integration task” designed by Moustafa et al. (23). In this task, subjects were presented a clock face, which had a single arm that made a full turn over the course of 5 sec. Participants were asked to press a button on a response pad at some point before the arm made a full turn. Participants were instructed to choose the button based on the current state of the clock face, in an attempt to maximize their utility (reward). The task was designed to assess participants’ ability to learn and apply a behavioral recipe for avolition and anhedonia. In light of previous studies suggesting that SZ is associated with abnormal DA signaling (high tonic DA but low phasic DA) and impaired D1 function in particular (25–26), that SZ patients have reduced prefrontal (and particularly, orbitofrontal) DA levels (22), we hypothesized that SZ patients would show not only differences in speeding and slowing as a function of prediction errors but also whether they would exhibit uncertainty-driven exploration.
Table 1. Demographic and Clinical Characteristics of Patients and CN

<table>
<thead>
<tr>
<th></th>
<th>SZ (n = 51)</th>
<th>CN (n = 39)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>42.77 (10.08)</td>
<td>43.49 (10.68)</td>
<td>.74</td>
</tr>
<tr>
<td>Education</td>
<td>12.80 (2.27)</td>
<td>14.89 (2.06)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Parental Education</td>
<td>13.56 (3.07)</td>
<td>13.26 (2.26)</td>
<td>.67</td>
</tr>
<tr>
<td>WASI Estimated Full-Scale IQ</td>
<td>96.86 (13.71)</td>
<td>113.13 (12.01)</td>
<td>.001</td>
</tr>
<tr>
<td>% Male</td>
<td>72.5%</td>
<td>66.7%</td>
<td>.35</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>56.9%</td>
<td>64.1%</td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>35.3%</td>
<td>35.9%</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2.0%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.0%</td>
<td>5.9%</td>
<td></td>
</tr>
<tr>
<td>Antipsychotic Medications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Conventional</td>
<td>4%</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>% Atypical</td>
<td>100%</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Clozapine</td>
<td>54%</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td>CPZ equivalent dosage</td>
<td>578 (394)</td>
<td>462 (371)</td>
<td>.24</td>
</tr>
<tr>
<td>BPRS Symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>2.28 (8.48)</td>
<td>1.38 (4.87)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Positive</td>
<td>2.50 (1.22)</td>
<td>2.48 (1.99)</td>
<td>.77</td>
</tr>
<tr>
<td>Disorganized</td>
<td>1.45 (0.71)</td>
<td>1.29 (0.47)</td>
<td>.35</td>
</tr>
<tr>
<td>Total</td>
<td>38.37 (9.30)</td>
<td>34.4 (6.70)</td>
<td>.09</td>
</tr>
</tbody>
</table>

SZ, schizophrenia; WASI, Wechsler Abbreviated Scale of Intelligence; CPZ, chlorpromazine.

Analyses conducted on high (HI-NEG) and low negative (LOW-NEG) symptom groups and healthy control (CN) groups. The most frequently used medication was clozapine, used alone (n = 21), in conjunction with risperidone (n = 6), or in conjunction with aripiprazole (n = 1). Risperidone prescribed alone (n = 6) or in conjunction with olanzapine (n = 2) was the second most frequently prescribed antipsychotic. Patients were also prescribed olanzapine (n = 8), fluphenazine (n = 1), ziprasidone (n = 2), or quetiapine (n = 3). One patient was prescribed haloperidol in conjunction with quetiapine. Mean Brief Psychiatric Rating Scale (BPRS) scores indicated that patients experienced a moderate level of symptom severity at the time of testing: Total (mean = 36.34, SD = 8.26); Positive (mean = 2.42, SD = 1.10); Negative (mean = 1.73, SD = .71); Disorganized (mean = 1.41, SD = .51).

full turn. After each response, participants were informed whether they had won points and, if so, how many. The trial ended once the subject responded with the game pad or if the 5-sec duration elapsed and the subject did not respond. The intertrial-interval was set at 1 sec. Participants completed four separate conditions, each consisting of 50 trials, in which reward probability and magnitude varied as a function of time elapsed on the clock until response. In the three primary conditions (DEV, CEV, and IEV), the number of points (reward magnitude) increased, whereas the probability of receiving the reward decreased over time within each trial. Feedback was presented on the screen in the format of “You win XX points!” Functions within each condition were designed such that the expected value (probability × magnitude) decreased (DEV), increased (IEV), or remained constant (CEV), across the 5-sec trial duration (Figure 1). Thus, in the IEV condition, early responses yielded a small number of points (lower than expected on average), and the associated negative prediction errors should lead to NoGo learning and slowed responses. In contrast, early responses in the DEV condition yielded a higher number of points than expected and should therefore lead to Go learning/speeding. Slower responses in the IEV condition yielded more points on average, whereas in the DEV condition faster responses yielded more points.

We included, in addition to these primary conditions, a condition where expected value remains constant (like CEV) but reward probability increases and magnitude decreases as time elapses on the clock (i.e., the opposite to CEV). Because both CEV and CEV reversed (CEVR) have equal expected values across the entire clock face, any difference in response time in these two conditions can be attributed to potential bias of a participant to learn more about reward probability than about magnitude or vice-versa. Specifically, if a subject waits longer to respond in CEVR than in CEV, it can be inferred that the participant is risk averse, because they value higher probabilities of reward more than higher magnitudes of reward.

Order of condition (CEV, DEV, IEV, CEVR) was counterbalanced across participants, and a rest break was given between each of the conditions (i.e., after every 50 trials). At the beginning of each condition, subjects were instructed to respond at different times to find the timer on the clock that would allow them to win the most points; however, they were not told about the different rules for each cond-
speed up or slow down and use positive and negative prediction errors across all conditions. The parameter was used to test the possibility that individuals with SZ have a reduced tendency to appropriately explore alternative actions in the hope that they might produce better outcomes.

Data Analysis

Behavioral analyses examined RT for each condition, either for the entire block or the difference score between the second and first half of trials in each condition as indicated in the text. Repeated measures analyses of variance (ANOVA), one-way ANOVAs, t tests, and χ2 analyses were calculated to determine group differences. Spearman correlations were calculated to examine relationships between test data and symptoms. The Greenhouse-Geisser correction was applied in instances when the assumptions of sphericity or covariance were violated. Scheffe contrasts were additionally performed as post hoc tests. Wilcoxon–Mann–Whitney tests were used to examine group differences on modeling parameters. Initial analyses examined between-group differences in patients and CN. However, given that SANS scores are typically bimodally distributed, we examined the role of negative symptoms with between-group analyses (i.e., comparing high negative symptom, low negative symptom, and control groups) but also reported correlations for completeness. Data were analyzed with SPSS version 12 software (SPSS, Chicago, Illinois).

Results

Go Versus NoGo Learning and Uncertainty-Driven Exploration

Analysis of behavioral data indicated that in both SZ patients and CN, RTs in the IEV condition (NoGo learning) were significantly slower than the DEV condition (Go learning) (CN: t = −4.48, p < .001; SZ: t = −4.99, p < .001), suggesting that both groups learn to adapt RTs in the expected direction (Table S1 in Supplement 1). However, these overall means calculated across the entire block of trials mask differences in learning from the beginning to the end of the condition. As such, difference scores were computed separately for each condition to estimate RT adaptation from the first half of trials to the second half of trials (second half of trials — first half of trials). Consistent with hypotheses, SZ patients failed to learn to speed up by the end of the block in the DEV (Go learning) condition as much as CN but perform similarly to CN in the IEV (NoGo) and CEV (Control) conditions (Figure 2). This was confirmed statistically by separate repeated measures ANOVAs, which indicated that groups significantly differed on the DEV condition \( F(1,88) = 9.49, p = .003 \) \((\eta^2 = .10)\) but not the IEV \( F(1,88) = .01, p = .913 \) \((\eta^2 = .00)\) or CEV conditions \( F(1,88) = 1.86, p = .176 \) \((\eta^2 = .02)\).

Table 2. Reinforcement Learning Domains Assessed by Computational Modeling Parameters

<table>
<thead>
<tr>
<th>Modeling Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>Estimates baseline motor response tendency independent of other factors.</td>
</tr>
<tr>
<td>λ</td>
<td>A response recency parameter scaling the impact of the previous response’s RT on the current choice, independent of any change in value.</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>Predicts trial-by-trial RT swings to occur when one is relatively more uncertain about the reward probabilities for fast or slow responses. Thus, with sufficiently high values, RT swings are predicted to occur in the direction of greater uncertainty about the likelihood that outcomes might be better than the status quo.</td>
</tr>
<tr>
<td>( \alpha G )</td>
<td>The degree to which individuals speed RTs as a function of positive prediction errors.</td>
</tr>
<tr>
<td>( \alpha N )</td>
<td>The degree to which individuals slow RTs as a function of negative prediction errors.</td>
</tr>
<tr>
<td>( \rho )</td>
<td>Predicts the extent to which individuals adjust RTs in the direction of greater probability of obtaining a positive outcome on the basis of the observed reward statistics.</td>
</tr>
<tr>
<td>( \nu )</td>
<td>A “going for gold” parameter, which predicts that participants will adjust RTs toward that which has produced the single largest reward experienced thus far.</td>
</tr>
</tbody>
</table>

RT, reaction time.
HI-NEG group showed significantly reduced speeding from the first to second half of trials for CEV, DEV, and IEV conditions in schizophrenia (SZ) patients and healthy control subjects (CN). **SZ** = white bars; **CN** = purple bars. Values reflect the mean RT change from beginning to end of the block for each condition (i.e., the relative learning within that condition). More negative values reflect that subjects learn to speed responses (Go), and more positive values reflect learning to slow down to obtain rewards within that condition (NoGo). Other abbreviations as in Figure 1.

These behavioral findings are consistent with intact NoGo learning in **SZ**. Computational modeling was used to obtain a richer understanding of these behavioral findings on Go and NoGo learning. Overall, the computational model encompassing the specified combination of parameters (Results in Supplement 1) and the best fit to the data in our previous study also provided a reasonable fit to the behavioral data here (Figure 3).

Significant parameter differences between **SZ** patients and CN were observed for \( \alpha \), the degree to which exploration occurs in proportion to relative uncertainty about reward outcomes \( F(1,88) = 9.1, p = .003 \). These differences remain significant after Bonferroni correction (Figure 4). Additional analyses also confirmed that the exploration effect in **SZ** was specific to uncertainty, because groups did not differ in measures of overall RT variability or RT swings (Figure S4 in Supplement 1). There was also a trend for \( \alpha G \) to be smaller in patients (Wilcoxon–Mann–Whitney test, two-tailed \( p = .07 \)), consistent with the behavioral results, whereby patients exhibited deficits in learning to speed responses in the DEV condition. A follow-up logistic regression with both parameters entered as predictors confirmed that both the explore \( (p < .02) \) and \( \alpha G \) parameters \( (p = .028) \) were independently predictive of \( \text{SZ} \). There were no other significant differences between patients and CN in any of the other parameters (Table 3).

A regression analysis shed further light onto this interpretation, revealing that individual differences in the tendency to speed up to maximize rewards in the DEV condition is predicted by \( \alpha G \) \( (p = .018) \), such that higher parameters were associated with increased speeding. No such difference was seen in terms of the model parameter estimating the degree to which participants slow down as a function of negative prediction errors.

**Negative Symptoms**

We also conducted analyses examining behavioral and modeling parameters in patients with high negative symptoms (HI-NEG), low negative symptoms (LOW-NEG), and CN. As can be seen in Figure 5, the HI-NEG group showed significantly reduced speeding from the first to second half of trials in the DEV condition, consistent with a Go learning deficit in HI-NEG patients. One-way ANOVAs, conducted with difference scores (second half of trials — first half of trials) as the dependent variable, support this interpretation, indicating that the three groups significantly differed on DEV \( F(2,85) = 4.78, p = .01 \) \( (\eta_2 = .11) \) but not IEV \( F(2,85) = .37, p = .73 \) \( (\eta_2 = .01) \) or CEV \( F(2,85) = 99, p = .73 \) \( (\eta_2 = .02) \). Post hoc Scheffe contrasts conducted for the DEV change condition were significant between the HI-NEG and CN \( (p = .02) \) but not the LOW-NEG and CN \( (p = .14) \) or HI-NEG and LOW-NEG \( (p = .68) \) groups. The correlation between the SANS total score and DEV and IEV conditions was nonsignificant. The discrepancy between the significant between-subjects analysis on the DEV condition and nonsignificant correlation between negative symptoms and DEV learning is likely because Go learning deficits were most pronounced in HI-NEG patients.

Figure 2. Mean reaction time (RT) difference score from first half of trials to second half of trials for CEV, DEV, and IEV conditions in schizophrenia (SZ) patients and healthy control subjects (CN). SZ = white bars; CN = purple bars. Values reflect the mean RT change from beginning to end of the block for each condition (i.e., the relative learning within that condition). More negative values reflect that subjects learn to speed responses (Go), and more positive values reflect learning to slow down to obtain rewards within that condition (NoGo). Other abbreviations as in Figure 1.

Figure 3. Response times as a function of trial in all 90 subjects (A) and computational model fits (B). The figures depict response times as a function of trial number, smoothed (with weighted linear least squares fit) over a 10-trial window in (A) all 90 participants, (B) computational model fits. Overall, relative to baseline CEV response times, participants speed up in DEV and slow down in IEV. The CEVR response times are also slowed in both data and model due to a high frequency of negative prediction errors for early responses (see Moustafa et al. [23] and Frank et al. [32]). Abbreviations as in Figures 1 and 2.

Figure 4. Uncertainty-driven exploration in individuals with schizophrenia (SZ) and healthy control subjects (CN). The explore parameter estimated from the model is reduced in schizophrenia (**\( p < .01 \)).
patients but still present in LOW-NEG patients as well, thereby attenuating the strength of the correlation.

In a separate analysis of behavioral data, HI-NEG patients also failed to show either a probability or magnitude bias, whereas CN and LOW-NEG both showed a bias to learn more about probability than magnitude (Figure S1 in Supplement 1 for discussion).

One-way ANOVAs indicated that the three groups failed to significantly differ on any parameter other than exploration [F(2,85) = 5.02, p = .009] (HI-NEG: mean = 1187, SD = 1561; LOW-NEG: mean = 1323, SD = 1678). Post hoc Scheffe contrasts indicated significant differences between HI-NEG and CN (p = .03) subjects; however, LOW-NEG and CN (p = .06) and HI-NEG and LOW-NEG (p = .97) did not significantly differ. Interestingly, correlational analyses indicated that the dramatic reduction in exploration was most severe in patients with high avolition-anhedonia SANS summary scores (r = −.28, p < .05). There were no significant correlations between ε and the restricted affect (SANS alogia + blunted affect items) summary score (r = .05) or the SANS total, suggesting that the relationship might be specific to the avolition-anhedonia domain. Follow-up correlational analyses with the avolition and anhedonia global scores indicated that the relationship with exploration was specific to anhedonia (anhedonia r = −.44, p < .01; avolition r = −.15, p > .3) (Figure 6). The test for significant differences between these correlations approached significance (z = −1.54, p = .06).

Given the unique association with anhedonia, we further investigated whether the association between anhedonia and exploration was specific to uncertainty and that anhedonia was only associated with uncertainty-driven exploration and not overall RT variability or consecutive RT variance. Furthermore, control model simulations revealed that other models of RT swings, including parameters for lose-switch or regression to the mean did not correlate with anhedonia (Results and Figure S2 in Supplement 1).

**Antipsychotic Medication**

Correlational analyses indicated that CPZ dosage was not significantly correlated with behavioral performance in any of the conditions (all p values > .16) or modeling parameters (all p values > .14). Analyses examining between-group differences in patients categorized as a function of low- and high-potency D2 blockade antipsychotics indicated no differences between medication groups in behavioral task conditions (Results in Supplement 1).

**Discussion**

Two main findings emerged from the current study. First, behavioral data indicated that patients were less able to learn to speed up to maximize rewards, which is consistent with a Go learning deficit. The model simulations suggest that this deficit might at least in part be due to lower αG parameter, because a regression analysis revealed that individual differences in the tendency to speed up to maximize rewards in the DEV condition is predicted by αG, such that higher parameters were associated with increased speeding. Given that SZ showed a deficit in both αG and the DEV but not αN or IEV, we feel that the results of the computational model provide further confidence that the deficits specific to Go learning in SZ are reliable. Furthermore, symptom subgroup analyses revealed that, in terms of DEV performance, Go learning deficits are most severe in patients exhibiting greater severity of negative symptoms. These findings are consistent with our previous probabilistic selection study indicating that SZ is associated with impaired Go learning and intact NoGo learning (19). When viewed in conjunction with neurocomputational models of corticostriatal circuitry in reinforcement learning (7,8), the current behavioral and modeling findings are suggestive of potential dysfunction in the direct D1-driven BG pathway leading to abnormalities in using positive feedback to guide behavior, with relatively intact function in the D2-driven, indirect pathway leading to normal use of probabilistic negative feedback in decision making. This BG-based account is supported by other evidence indicating

![Figure 5](https://www.sobp.org/journal)

**Figure 5.** Go Learning high (HI-NEG) and low (LOW-NEG) negative symptom patient groups, and CN subjects. Mean RT change from beginning to end of block for the DEV condition (Go Learning). More negative values reflect better Go Learning (i.e., learning to speed up from the first half of trials to the second half of trials). Other abbreviations as in Figure 2.

![Figure 6](https://www.sobp.org/journal)

**Figure 6.** Uncertainty-driven exploration as a function of anhedonia. Left: scatter plot across all patients. Right: means for each level of anhedonia. Error bars reflect SEM.
that BG DA acts to speed responding toward rewarding cues (38,39) as well as pharmacological and animal studies showing that this process likely relies on D1-driven activation and Go learning (40–42). However, this interpretation is of course speculative and cannot be confirmed without conducting a study on unmedicated first-episode patients to see whether NoGo learning improves when patients are treated with D2-blocking antipsychotics.

A second major finding was that SZ patients exhibited a large and reliable reduction in the tendency to make exploratory behavioral adjustments toward responses that could potentially yield larger expected values than those obtained by staying with the status quo. Additionally, given that there was no association between anhedonia and overall RT variability or consecutive variance, anhedonia seems to be selectively associated with the failure to initiate the proactive strategy of adjusting responses to gather more information to reduce uncertainty about potential benefits of alternative behaviors. These findings demonstrate the usefulness of computational modeling approaches to psychiatry (43–47).

We posit that these effects are related to degradations in prefrontal cortical DA function, often attributed as a source of negative symptoms (24,26,48,49). This interpretation is supported by our recently reported gene-dose effect of the val/met polymorphism of the COMT gene in healthy individuals performing this same task (32), which indicated that the val/val genotype was characterized by the lowest degree of uncertainty-driven exploration and the met/met genotype with the greatest degree of exploration. Variations in COMT affect prefrontal and particularly orbitofrontal DA levels (22), and a recent study reported a COMT gene dose effect on orbitofrontal activity during reward receipt (50). Thus, together, these studies support the assertion that the val/val genotype shares features of cognitive dysfunction observed in SZ (51). Finally, ongoing imaging work in healthy individuals (N. Long, B.S., unpublished data, September 1, 2010), together with other related studies (31,52), suggest that relative uncertainty computations associated with exploration are represented in prefrontal cortical activation patterns. Finally, even if the computations of expected reward values are relatively intact in SZ, it is possible that patients with anhedonia explicitly assign a negative expected value to uncertain outcomes, due to their prior expectations (see Huys and Dayan [47] for a related model of depression). Regardless of the neural mechanism, our findings suggest that anhedonia might result from an inability to determine when to explore actions that might improve one’s ability to obtain rewards.

Of particular interest was that reduced uncertainty-driven exploration correlated with the avolition-anhedonia domain on the SANS but not the Restricted Affect factor. Additionally, the effect was more highly related to anhedonia than avolition. This result is potentially informative about differences in the pathology of these symptom domains. As rated by the SANS, anhedonia reflects a behavioral component of reward seeking (e.g., initiating social activities, sexual interest/and or activity, pursuing recreational activities, number of close relationships) rather than the capacity to experience pleasure, which is often inferred from behavior. Avolition items on the SANS are less related to reward-seeking behavior and more broadly related to the frequency with which patients initiate and persist in many kinds of tasks, which is likely to be influenced by a number of factors, such as disorganization, generalized cognitive impairment, and sedation. The significant correlation with anhedonia but not avolition might therefore reflect that reduced reward-seeking behavior in SZ is critically related to the extent to which patients make exploratory choices when they are uncertain about the value of alternative actions and whether they might produce better outcomes than the status quo.

Results should be viewed with certain limitations in mind. First, analyses regarding the role of medication on task performance should be viewed with caution, because CPZ equivalents for atypical medications might not be appropriate and D2 potency classifications provide only a gross estimate of the effects of different antipsychotics. A more definitive test of antipsychotic effects should be conducted in first episode patients tested on and off medications. Second, we did not collect DNA in this study, and it is unclear whether the COMT genetic effect observed in healthy individuals on exploration might partially contribute to the effects of anhedonia and SZ reported here. Finally, although the SANS is still the gold standard negative symptom assessment in the field, it has recently been suggested that newer measures being developed in response to the National Institute of Mental Health Measurement and Treatment Research to Improve Cognition in Schizophrenia (e.g., Kirkpatrick et al. [53]) initiative might provide a more comprehensive and current assessment of negative symptom dimensions. As such, it is unclear whether the relationship reported between SANS anhedonia and exploration might actually reflect some other component of negative symptoms on these newer scales.

In summary, the current findings have important implications for understanding the etiology of SZ. Results from the computational model and behavioral data indicate that patients have deficits in Go learning, which seem to be due to reduced sensitivity to positive prediction errors. Thus, patients show a reduced sensitivity to the impact of rewarding outcomes on future behavioral choices. Furthermore, patients display reduced uncertainty-driven exploration, which was specifically associated with greater severity of anhedonia. Thus, patients are less likely to explore and therefore less likely to discover that an alternative response might yield more rewarding outcomes. Although these deficits are independent of one another in the model, at a clinical level it is easy to imagine how these impairments might amplify one another and result in a narrow behavioral repertoire and a lack of goal-directed, reward-seeking behavior.

This research was supported by US National Institutes of Mental Health Grant R01 MH080066-01. We would also like to thank the subjects who participated in the study and staff at the Maryland Psychiatric Research Center who made the completion of the study possible. We are especially thankful to members of Dr. Gold’s laboratory, Jackie Kwanuka, Sharon August, Lindsay Phebus, Leeka Hubzin, and Tatyana Matveeva, who conducted subject recruitment and testing. We also thank Thomas Wiecki for helpful comments on an earlier draft of this manuscript.

All authors reported no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.


