Archival Report

Biological Psychiatry: CNNI

Impaired Expected Value Computations Coupled With Overreliance on Stimulus-Response Learning in Schizophrenia

Dennis Hernaus, James M. Gold, James A. Waltz, and Michael J. Frank

ABSTRACT

BACKGROUND: While many have emphasized impaired reward prediction error signaling in schizophrenia, multiple studies suggest that some decision-making deficits may arise from overreliance on stimulus-response systems together with a compromised ability to represent expected value. Guided by computational frameworks, we formulated and tested two scenarios in which maladaptive representations of expected value should be most evident, thereby delineating conditions that may evoke decision-making impairments in schizophrenia.

METHODS: In a modified reinforcement learning paradigm, 42 medicated people with schizophrenia and 36 healthy volunteers learned to select the most frequently rewarded option in a 75-25 pair: once when presented with a more deterministic (90-10) pair and once when presented with a more probabilistic (60-40) pair. Novel and old combinations of choice options were presented in a subsequent transfer phase. Computational modeling was employed to elucidate contributions from stimulus-response systems (actor–critic) and expected value (Q-learning).

RESULTS: People with schizophrenia showed robust performance impairments with increasing value difference between two competing options, which strongly correlated with decreased contributions from expected value-based learning (Q-learning). Moreover, a subtle yet consistent contextual choice bias for the probabilistic 75 option was present in people with schizophrenia, which could be accounted for by a context-dependent reward prediction error in the actor–critic.

CONCLUSIONS: We provide evidence that decision-making impairments in schizophrenia increase monotonically with demands placed on expected value computations. A contextual choice bias is consistent with overreliance on stimulus-response learning, which may signify a deficit secondary to the maladaptive representation of expected value. These results shed new light on conditions under which decision-making impairments may arise.

Keywords: Computational psychiatry, Decision making, Expected value, Motivational deficits, Reinforcement learning, Schizophrenia

https://doi.org/10.1016/j.bpsc.2018.03.014

Reinforcement learning (RL) and decision-making impairments are a recurrent phenomenon in people with schizophrenia (PSZ) and are thought to play a key role in abnormal belief formation (1) and motivational deficits (2). While many have emphasized an impairment in stimulus-response learning (3–5), multiple studies suggest that some of these deficits may in fact arise from overreliance on stimulus-response learning together with a compromised ability to represent the prospective value of an action or a choice (i.e., expected value) [e.g., (6,7); for an overview, see Waltz and Gold (2)]. However, such conclusions have typically been based on inferences rather than experimental designs intended to reveal such effects. Therefore, we formulated and tested two hitherto unexplored scenarios motivated by the posited computations under which deficits in the representation of expected value should be most evident.

Optimal decision making relies on a pas de deux between a flexible and precise representation of expected reward values, supported by orbitofrontal cortex (8–10), which is complemented by a gradual buildup of stimulus-response associations credited to dopaminergic teaching signals (reward prediction errors [RPEs]) that project to striatum (11,12). Previous work has demonstrated that maladaptive representations of expected value, rather than diminished stimulus-response learning per se, is one consistent feature of RL deficits in PSZ (13–16).

Findings of impaired representations of expected value in PSZ have often relied on computational models of learning and decision making. In RL computational frameworks, it is thought that Q-learning and actor–critic models capture expected value and stimulus-response learning, respectively. In Q-learning (17), the RPE (the difference between expectation and outcome) directly updates the expected value of a choice option—similar to the representation of a reward value by orbitofrontal cortex (18,19)—and response tendencies are driven by large action values. In contrast, in the actor–critic

ISSN: 2451-9022

framework (20), RPEs signaled by the critic—who observes outcomes—update its state value (e.g., being presented with a certain choice pair) rather than updating the value of each choice option separately. Importantly, the critic's RPE also updates the actor's response tendency for the chosen option. Thus, the actor develops response tendencies for choices associated with more positive (better than expected) than negative (worse than expected) RPEs signaled by the critic and not on the basis of an exact estimate of reward value. It is thought that the slow buildup of the actor's response tendencies, on the basis of an accumulation of RPEs, reflects dopamine-mediated changes in synaptic weights in basal ganglia (21–23). Crucially, because the RPE fulfills different roles in these two computational frameworks (i.e., updating reward value directly vs. modifying stimulus-response weights), it follows that, by definition, reward value is more precisely represented in the Q-learning framework than in the actor–critic framework. In one study, we showed that a computational modeling parameter that captured the balance between Q-learning and actor–critic-type learning was tilted in favor of the latter in PSZ, suggesting relative underuse of expected value and, perhaps secondarily, overreliance on stimulus-response learning (6). To date, however, little is known about the conditions under which deficits in the computation of expected value should be most observable.

Therefore, we sought to test two predictions of our theoretical account, which emphasizes maladaptive representation of expected value (Q-learning) in PSZ:

1. Counterintuitively, and in contrast to many situations in which PSZ may be most impaired at high levels of difficulty, our model based on less precise representations of reward value (decreased Q-learning) predicts that PSZ should suffer the largest decision-making deficits for the easiest value discriminations—that is, when the value difference between two competing options increases.
2. If the relative contribution of actor–critic-type learning is greater in PSZ—because of a decrease in Q-learning—then one might observe biases in action selection among choice options that have identical reinforcement probabilities, based on differences in critic RPEs. In the actor–critic architecture, RPEs are evaluated relative to the overall reward rate of the context. Thus, rewards presented in contexts with low reward rates elicit larger RPEs than those presented in more deterministic contexts. Therefore, a second diagnostic prediction is that PSZ should elicit observable context-dependent choice biases, even among items with identical reinforcement histories.

In the current study, we tested these two hypothesized consequences of deficits in the representation of expected value using a modified RL paradigm. Participants were presented with two pairs of stimuli with identical reward value; one pair was presented in a reward-rich context (where the other pair had a higher reward rate), while the other pair was presented in a reward-poor context (where the other pair had a lower reward rate). Afterward, participants were presented with old and novel combinations of choice options. We exploited the wide range in reward value to test our hypothesis relating to performance deficits as a function of the value difference between two competing options. Pairs with identical reward value in different contexts allowed us to address hypotheses relating to a contextual choice bias that should be present to the degree that individuals rely on actor–critic-type learning.

To accomplish these aims, we used a previously validated hybrid computational model that estimates one's tendency to use Q-learning versus actor–critic along a parametric continuum (6). As observed previously (6), we expected PSZ to rely less on Q-learning than on actor–critic, resulting in the aforementioned deficits.

METHODS AND MATERIALS

Sample

We recruited 44 participants with a DSM-IV diagnosis of schizophrenia or schizoaffective disorder and 36 healthy volunteers (HVs). Of these participants, 2 PSZ were excluded—1 who was mistakenly administered an old version of the task and another participant who consistently performed far below chance—leaving a sample of 42 PSZ. PSZ were recruited through clinics at the Maryland Psychiatric Research Center. HVs were recruited by advertisements posted on the Internet (Craigslist) and via notices on bulletin boards in local libraries and businesses. A diagnosis of schizophrenia or schizoaffective disorder in PSZ, as well as the absence of a clinical disorder in HVs, was confirmed using the Structured Clinical Interview for DSM-IV Axis I Disorders (24). The absence of an Axis II personality disorder in HVs was confirmed using the Structured Interview for DSM-III-R Personality Disorders (25). In total, 37 PSZ were diagnosed with schizophrenia and 5 PSZ were diagnosed with a schizoaffective disorder. Comorbid disorders included obsessive-compulsive disorder (n = 1), anxiety disorder (n = 1), and a cannabis dependence disorder (in remission; n = 1). All PSZ were on a stable antipsychotic medication regimen. No changes in medication dose/type were made during the 4 weeks leading up to study participation. Major exclusion criteria included pregnancy, current illegal drug use, substance dependence (during past year), a neurological disorder, and/or a medical condition affecting study participation. All participants provided written informed consent. The study was approved by the Institutional Review Board of the University of Maryland School of Medicine.

Clinical Ratings

The avolition-apathy and asociality-anhedonia (AA) subscales of the Scale for the Assessment of Negative Symptoms (26) and the positive symptom subscale of the Brief Psychiatric Rating Scale (27) were used as measures of negative and positive symptoms, respectively. See the Supplement for details on these scales as well as on other sociodemographic and clinical variables.

RL Paradigm

Participants completed an RL paradigm consisting of a 320-trial learning phase and a 112-trial transfer phase.

Learning Phase. Participants were presented with pairs of stimuli and were asked to select one using their left (left choice) or right (right choice) index finger, after which they received

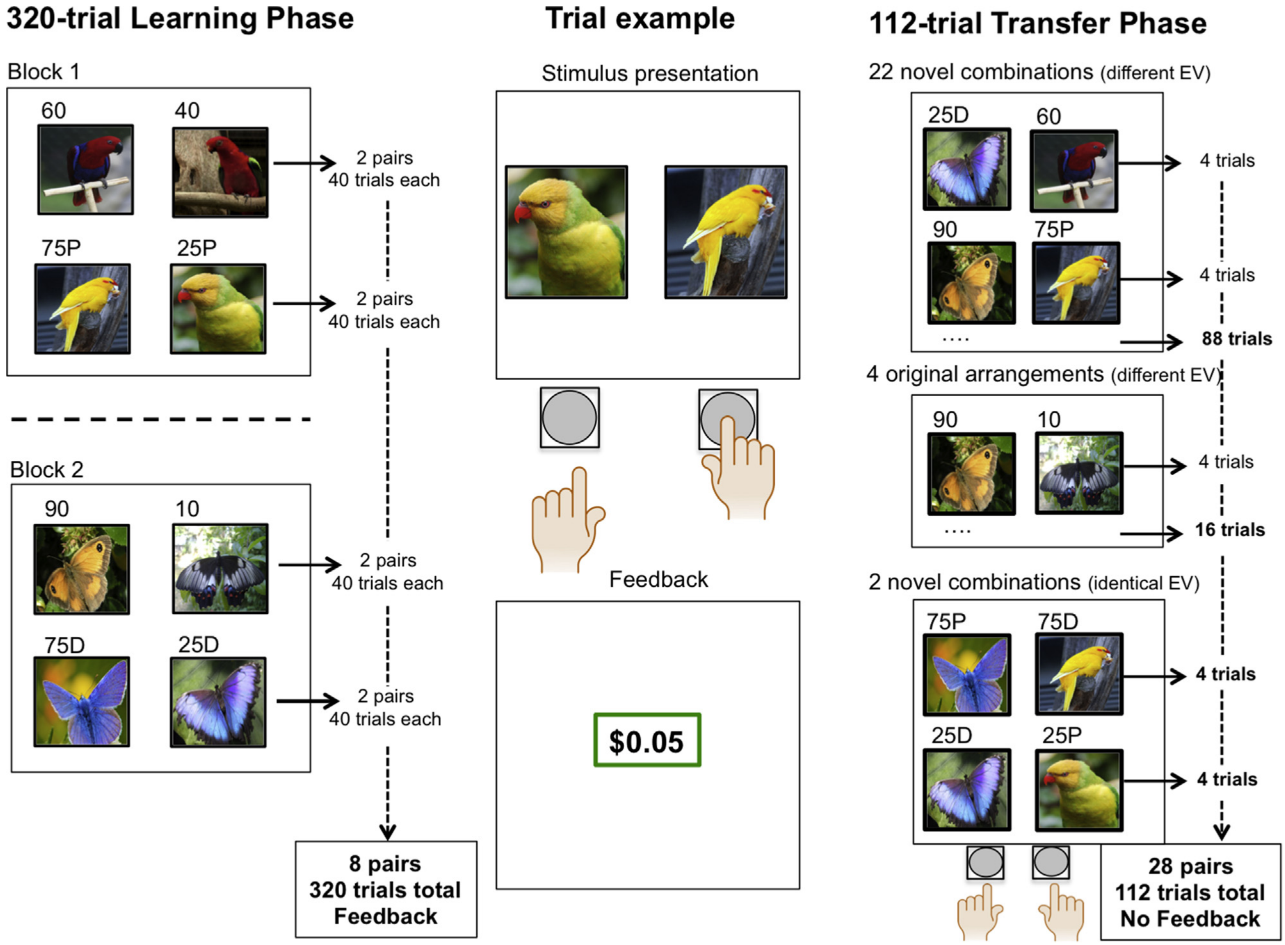


Figure 1. Schematic of the reinforcement learning paradigm. EV, expected value.

positive (+$.05) or neutral ($0.00) feedback (Figure 1). The learning phase consisted of two 160-trial blocks in which four pairs were presented per block. In one block, two pairs were presented that rewarded optimal choices 90% of the time (and no reward on the remaining 10% of trials) together with two pairs that rewarded optimal choices 75% of the time (no reward on 25%). In the other block, two pairs were presented that rewarded optimal choices 75% of the time (no reward on 25% of trials) together with two pairs that rewarded optimal choices 60% of the time (no reward on 40%) (Figure 1). Feedback contingencies for the suboptimal choice were the mirror image of the optimal choice. Trial presentation was pseudorandomized within each block, and block order was counterbalanced among participants, as were block theme (butterfly or bird stimuli), option-probability pairing, and option-position pairing (left/right side of screen). Note that two of four 75-25 pairs always comprised bird-themed stimuli, and the other two pairs always comprised butterfly-themed stimuli (Figure 1).

By combining 75-25 pairs with more deterministic (90-10) and probabilistic (60-40) pairs in separate blocks, we aimed to investigate context-dependent RL, meaning that perceived choice value (here, 75-25 pairs) might be dependent on contextual reward rate (the average reward rate of optimal choice options within a block). Henceforth, we refer to 75-25 pairs that were presented together with 90-10 pairs as 75-25D (with the letter D representing the more deterministic context) and refer to 75-25 pairs that were presented with 60-40 pairs as 75-25P (with the letter P representing the more probabilistic context).

Transfer Phase. The 112-trial transfer phase served two purposes: 1) to assess the ability to compare choice options using their reward value and 2) to provide a formal test of a contextual choice bias.

Every possible combination of two-choice options (new and original combinations) was presented, and the participant was instructed to “select the option that was rewarded most often” (Figure 1). To prevent further learning, no feedback was delivered. Combining all possible choice options yielded 28 combinations with nonidentical expected value, 22 novel combinations (e.g., 90-60, 75D-10) and 4 original combinations (90-10, 75-25D, 75-25P, and 60-40). In addition, we produced two novel combinations with identical expected value: 75P-75D and 25P-25D. Supplemental Table S1 provides an overview of all 28 transfer pairs.

Note that there were two pairs of each contingency in the learning phase, with the exception of 75-25 pairs, of which there four pairs. Thus, although every unique combination of choice options was presented only once in the transfer phase, there were always four presentations of each expected value combination; for example; with two 90-10 pairs from the learning phase, one can generate four unique 90-10 combinations in the transfer phase.

Computational Model

In an attempt to relate deficits in expected value and a contextual choice bias to latent variables, we used a previously validated hybrid model allowing for combined influences of Q-learning (action selection as a function of expected reward value) and actor–critic (basal ganglia–dependent stimulus-response learning) frameworks on decision making (6). This model was compared with a basic actor–critic and Q-learning model. The hybrid model had the best trade-off between model complexity, fit, and posterior predictive simulations, and it contained six free parameters—a critic learning rate (α_c), an actor learning rate (α_a), a Q-learning rate (α_q), an inverse temperature (β), mixing (m), and an undirected noise (ε) parameter—which were estimated for every subject via maximum likelihood optimization. The Supplement and Supplemental Figure S1 contain a detailed description of the model and selection procedure, including the use of context (i.e., block)-dependent state values for the critic and an ε-softmax choice function. After fitting the hybrid model to the learning phase data, the final action weights of all eight original pairs were used to simulate transfer phase performance for all pairs (n [simulations] = 250 for every participant).

Statistical Analyses

Performance on both pairs of each probability level (90-10, 75-25D, 75-25P, and 60-40) was averaged. Next, learning phase trials were divided into four bins of 10 trials for each probability level. A 2 × 4 × 4 repeated-measures analysis of variance using group status (predictor) and probability level (four levels) as predictors and trial bin (bins; four levels) as dependent variables was run to test for a group by condition by time interaction. Group by time and group by condition interactions were also investigated. Greenhouse-Geisser sphericity-corrected values were reported when assumptions were violated.

Transfer phase accuracy was averaged across all four presentations of every unique combination (n = 28) of expected values and compared using two-sample t tests. Transfer phase pairs were next ranked on their value difference (see Supplemental Table S1 for details regarding trial combinations). A logistic regression analysis with value difference (left–right option) as predictor and correct choice (left vs. right button) as dependent variable was conducted to test the hypothesis that PSZ show impaired performance with increasing value difference. Individual value difference slopes were compared in a two-sample t test.

Context-dependent learning was investigated using a two-sample t test as well as a one-sample t test to compare preference for either option against chance. As a direct measure of context-dependent learning, we focused on trials where 75P was coupled with 75D. As an indirect measure of context-dependent learning, average performance on all trials where 75P and 75D stimuli were presented with any other option (excluding the 25 stimulus with which they were originally partnered) was compared in a 2 × 2 group by pair analysis of variance. Supplemental Table S1 provides a detailed overview of transfer phase trials that were used for this analysis. Significance thresholds for performance data were set to $p < .05$.

Correlation analyses among model parameters, clinical variables, and psychometric variables were carried out using Pearson's r, Spearman's ρ, and subgroup splits (the latter two when distributions were skewed). Significance thresholds for correlation/subgroup split analyses were Bonferroni corrected for the number of parameters in the model ($p_{\text{Bonferroni corrected}} = .05$; $p_{\text{uncorrected}} = .008$).

RESULTS

Demographics

Participant groups were matched on most demographics. However, PSZ did have a lower IQ score, as well as poorer Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery performance, than HVs (Table 1).

Learning Phase Performance

We observed a group × probability interaction ($F_{3,228} = 4.39$, $p = .005$), such that HVs outperformed PSZ in the 90-10 ($p = .002$), 75-25D ($p = .007$), and 75-25P ($p = .04$) probability conditions but not in the 60-40 ($p = .63$) probability condition (Figure 2A). Group × probability × time ($F_{9,684} = 0.98$, $p = .46$) and group × time ($F_{3,228} = 0.62$, $p = .60$) interactions were not significant. Performance on 60-40 trials in bin 4 was significantly above chance for both groups (HVs: $t_{35} = 2.88$, $p = .007$; PSZ: $t_{41} = 2.64$, $p = .01$). See Supplemental Figure S2 for individual data points for each probability level.

Transfer Phase Performance

Despite poorer learning accuracy in PSZ, there were no group differences in transfer accuracy for 90-10, 75-25D, 75-25P, or 60-40 pairings (all $ps > .39$) (Figure 2B), with accuracy above chance on all pairs.

Smaller Performance Improvements With Increasing Value Difference in PSZ

Accuracy on all novel pairs is shown in Supplemental Figure S3. When all combinations of reward contingencies were considered, accuracy on trials with a value difference of 35 ($t_{76} = 3.55$, $p = .06$), 50 ($t_{76} = 4.26$, $p = .04$), and 60 ($t_{76} = 4.08$, $p = .05$) were (trendwise) greater in HVs compared with PSZ (Figure 2C). This was also true when using only novel pairs or only pairs consisting of one choice option from each context (Figure 2C). To formally test the presence of a greater accuracy deficit with increasing value difference, we compared individual slopes from a logistic regression predicting accuracy as a function of value difference. Using all pairs ($t_{74} = 5.84$, $p = .02$), novel pairs ($t_{74} = 6.99$, $p = .01$), and novel context pairs ($t_{73} = 6.05$, $p = .02$), the slope for HVs was always greater than that for PSZ (these results could not be used in 2–4 participants

Table 1. Demographics

	HVs (n = 36)	PSZ (n = 42)	t or χ^2	p
Age, Years	42.81 (8.86)	44.60 (8.26)	−0.92	.36
Gender, Female/Male	12/24	13/29	0.05	.82
Race, African American/ Caucasian/Other	11/24/1	13/25/4	2.74	.60
Education Level, Years	14.86 (1.99)	12.69 (2.20)	4.49	< .001
Maternal education level[a]	13.60 (2.19)	13.46 (2.51)	0.25	.80
Paternal education level[b]	13.29 (3.05)	13.89 (4.20)	−0.70	.48
WASI-II IQ Score	114.86 (10.59)	98.10 (14.89)	5.76	< .001
MATRICS Domains[c]				
Processing speed	54.66 (9.47)	35.12 (11.57)	7.99	< .001
Attention/vigilance	51.77 (11.47)	41.45 (12.44)	3.75	< .001
Working memory	54.23 (10.16)	38.02 (11.13)	6.62	< .001
Verbal learning	50.11 (10.58)	36.69 (8.10)	6.30	< .001
Visual learning	45.46 (11.23)	35.02 (13.49)	3.64	< .001
Reasoning	53.84 (9.99)	43.02 (9.64)	4.82	< .001
Social cognition	50.91 (8.93)	36.83 (11.12)	6.04	< .001
Antipsychotic Medication[d]				
Total chlorpromazine	–	332.36 (424.21)	–	–
Total haloperidol	–	6.88 (9.10)	–	–
Clinical Ratings				
BPRS positive (sum)	–	9.30 (5.37)	–	–
SANS AA/RF (sum)	–	17.00 (7.73)	–	–
SANS AFB/Alog (sum)	–	10.67 (7.89)	–	–

Values are presented as mean (SD) or n.

AA/RF, avolition-apathy (including current role and function) and asociality-anhedonia sum scores; AFB/Alog, affective flattening and alogia sum scores; BPRS, Brief Psychiatric Rating Scale; HVs, healthy volunteers; MATRICS, Measurement and Treatment Research to Improve Cognition in Schizophrenia; PSZ, people with schizophrenia; SANS, Scale for the Assessment of Negative Symptoms; WASI-II, Wechsler Abbreviated Scales of Intelligence–Second Edition.

[a] Maternal education missing for 1 HV and 5 PSZ.

[b] Paternal education missing for 1 HV and 4 PSZ.

[c] MATRICS ratings and IQ score missing for 1 HV.

[d] Chlorpromazine and haloperidol missing for 1 PSZ.

due to limited choice variability) (Figure 2D). This all suggests that PSZ, compared with HVs, improved less as the value difference between two competing stimuli increased, thereby confirming our initial hypothesis.

Context Influences Perceived Choice Value in PSZ but Not in HVs

A direct comparison of 75D-75P performance revealed no significant group difference (t_{76} = 1.61, p = .21) (Figure 2E). However, PSZ (one-sample t test against chance: t_{41} = −2.10, p = .04), but not HVs (t_{35} = 0.01, p = .99), did show a significant preference for 75P over 75D. The more indirect group × pair interaction for 75P and 75D performance versus other options showed similar numerical patterns but was not significant ($t_{1,67}$ = 2.11, p = .15). Nevertheless, PSZ (t_{41} = −2.52, p = .015), but not HVs (t_{35} = −0.67, p = .51), more often selected 75P than 75D when paired with another option (Figure 2E). The direct and indirect measures of context sensitivity correlated in PSZ (Pearson's r = −.53, p < .001). Taken together, these results provide subtle yet consistent evidence that context may affect perceived choice value in PSZ but not in HVs. To formally test whether the trial by trial pattern of choices can be explained by context-dependent value learning, we next turn to computational modeling results.

Computational Modeling

Model Parameters. Hybrid model parameters for HVs and PSZ are shown in Figure 3A and summarized in Table 2. The average mixing parameter was greater than .5 in PSZ and HVs, suggesting that both groups made more use of Q-learning compared with actor–critic-type learning. Importantly, as predicted, the m parameter was significantly greater in HVs than in PSZ (Table 2). This result points to a decrease in Q-learning and a relative increase in actor–critic-type learning in PSZ compared with HVs. In addition, the undirected noise parameter was greater in PSZ than in HVs (Table 2). See Supplemental Table S2 for individual parameter estimates.

Hybrid Model Simulations: Learning Phase. True to the actual learning phase data, model simulations revealed numerically greater performance in HVs relative to PSZ for 90-10, 75-25D, and 75-25P contingencies but not for 60-40 contingencies, which became (trend) significant when increasing the number of simulations [n(simulations) = 1000 shown in Figure 3B].

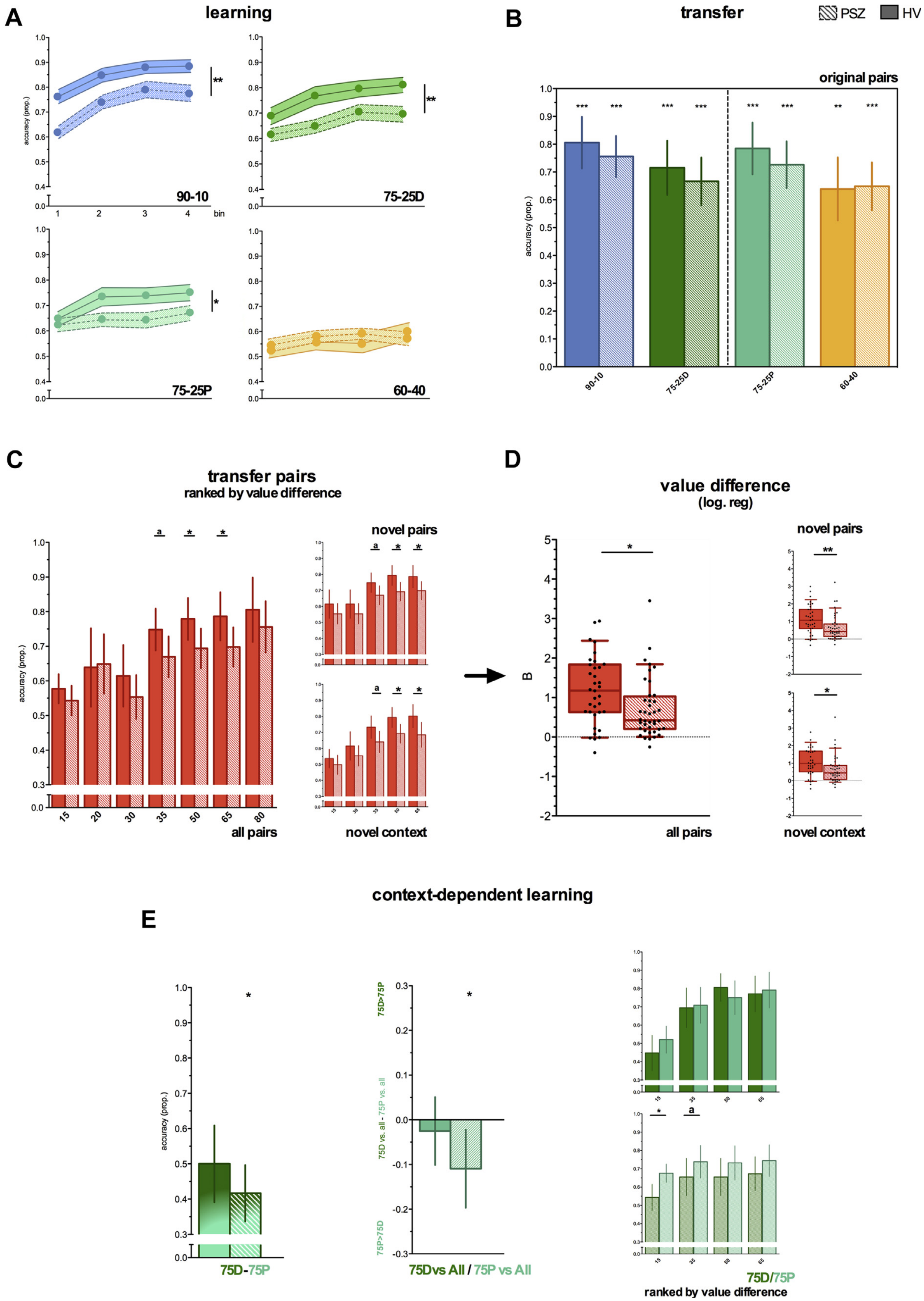
Hybrid Model Simulations: Transfer Phase. Given the low number of transfer phase trials for every combination (n = 4), and because the amount of undirected noise may be greater during learning compared with transfer phase performance, we set ε to 50% of the original value during transfer phase simulations. All findings remained when simulating transfer data with ε set to 100% (Supplemental Figure S4).

Simulated group differences in transfer phase accuracy on 90-10 (t_{76} = 1.93, p = .06), 75-25D (t_{76} = 2.53, p = .01), 75-25P (t_{76} = 2.30, p = .02), and 60-40 (t_{76} = 0.52, p = .60) pairs were subtle (Figure 3C), as was the case in the original data, yet (trended) significant for some pairs owing to the number of simulations [n(simulations) for all transfer data = 250] (Figure 3C). Importantly, simulated data from the hybrid model predicted numerically greater performance deficits in PSZ with increasing value difference (Figure 3D).

Hybrid Model Simulations: Context-Dependent Learning. The direct and indirect context effects in PSZ both were present in the simulated data (Figure 3E), that is, 1) a preference for 75P over 75D (t_{40} = 2.59, p = .01) and 2) better performance when 75P was paired with other choice options compared with when 75D was paired with other choice options (t_{39} = 2.03, p = .05). One outlier in the PSZ sample with high values overall/difference scores was removed from the simulated data; excluding this subject from the actual data did not change the results.

Evidence That Model Parameters Capture Task Performance

The m (Spearman's ρ = −.67, p < .001) and ε (Spearman's ρ = .38, p < .001) parameters significantly correlated with the slope of the value difference effect in the entire sample, suggesting that decreased reliance on Q-learning and greater


A
learning
accuracy (prop.)
90-10
75-25D
75-25P
60-40
bin
**
**
*
B
transfer
▨ PSZ
■ HV
original pairs
accuracy (prop.)
90-10
75-25D
75-25P
60-40
C
transfer pairs
ranked by value difference
accuracy (prop.)
15
20
30
35
50
65
80
all pairs
novel pairs
novel context
a
*
D
value difference
(log. reg)
B
all pairs
novel pairs
novel context
**
*
context-dependent learning
E
accuracy (prop.)
75D-75P
75D>75P
75D vs. all - 75P vs. all
75P>75D
75Dvs All / 75P vs All
75D/75P
ranked by value difference

undirected noise were associated with smaller performance improvements with increasing value difference (Figure 4A, B).

Next, we focused on the α_c parameter, which can produce the context effect within the actor–critic model. To demonstrate this, both m and ε were fixed to 0 and the direct and indirect context effects were simulated. This analysis removes contributions from Q-learning and undirected noise, while all other parameters were set to their original values. In PSZ, α_c correlated with the size of the simulated direct (Spearman's $\rho = -.42$, $p < .005$) and indirect (Spearman's $\rho = .47$, $p < .001$) context effects. This confirms our intuition that varying levels of critic learning rate are sufficient to account for the context effect. Moreover, when simulating data using individual m and ε parameters, α_c-weighted [$\alpha_c \times (1 - \mathrm{m})$, i.e., the degree to which α_c could have produced a context effect] also significantly correlated with the simulated indirect context effect (Pearson's $r = .34$, $p = .02$), while the correlation with the simulated (Pearson's $r = -.28$, $p = .07$) and actual (Pearson's $r = -.13$, $p = .41$) direct context effects was in the expected direction but not significant. This provides evidence that greater α_c values can account for a context-dependent choice bias, although this also crucially depends on the degree to which participants rely on Q-learning and the amount of undirected noise.

Associations With Clinical and Demographic Variables

Parameter estimates for low (avo−) and high (avo+) motivational deficit subgroups (median AA sum score = 17; 19 avo−; 23 avo+) are shown in Supplemental Table S3 and Supplemental Figure S5. The avo+ compared with avo− showed a selective increase in α_c ($t_{40} = 2.84$, $p_{\text{Bonferroni corrected}} = .04$); the same trend was observed for avo+ versus HV ($t_{57} = 2.54$, $p_{\text{Bonferroni corrected}} = .06$). Given that α_c strongly correlated with measures of context-dependent RL, and in light of the lower m parameter in PSZ relative to HVs, these results suggest that avo+ were more sensitive to context-dependent RL if they relied strongly on actor–critic-type learning.

Focusing on model parameters that could explain group differences in task performance, m (HVs low vs. high IQ: $t_{33} = 1.37$, $p_{\text{uncorrected}} = .18$; PSZ low vs. high IQ: $t_{42} = 1.04$, $p_{\text{uncorrected}} = .31$) and α_c (HVs low vs. high IQ: $t_{33} = 0.50$, $p_{\text{uncorrected}} = .62$; PSZ low vs. high IQ: $t_{42} = 1.40$, $p_{\text{uncorrected}} = .18$) were not associated with IQ. In PSZ (low vs. high IQ: $t_{42} = 2.56$, $p_{\text{Bonferroni corrected}} = .09$), but not in HVs (low vs. high IQ: $t_{33} = 0.49$, $p_{\text{uncorrected}} = .63$), there was a trend of a lower IQ being associated with more undirected noise (i.e., greater ε).

Finally, m, ε, and α_c were not associated with haloperidol equivalents (all $p_{\text{uncorrected}} > .71$) or age ($p_{\text{uncorrected}} > .32$). Scale for the Assessment of Negative Symptoms AA sum scores were not associated with model fit (Spearman's $\rho = -.003$, $p_{\text{uncorrected}} = .98$).

Follow-up analyses in a subsample of participants who performed particularly well are reported in the Supplement.

DISCUSSION

Using theory-based predictions, our primary aim was to investigate two hypothesized RL and decision-making deficits that could result from relative underuse of expected value. As predicted, PSZ showed robust performance impairments as the difference in reward value between two choice options increased.

Moreover, we observed a subtle yet consistent contextual choice bias that was not present in HVs: when presented with two options of identical reward value (75D and 75P), or when these options were paired with options of other reward value, PSZ preferred the 75 option from the more probabilistic context (75P).

Performance deficits amplified at greater levels of value difference are diagnostic of a change in the choice function rather than a general learning impairment, which would typically manifest in the opposite manner, that is, worse performance for more difficult judgments. These results are particularly noteworthy because they further corroborate the notion that some learning and decision-making deficits in PSZ are associated with a highly selective deficit in the representation of expected value. A more general learning impairment, potentially via altered dopamine-dependent stimulus-response learning (1,3,5), would predict performance impairments with increasing levels of difficulty. We have previously observed a hint for performance deficits at greater levels of value difference in other RL tasks (6,28), suggesting that this is a recurrent impairment in PSZ. Our computational model provides evidence that such impairments stem from a decrease in action value learning (Q-learning; via the m parameter) and a greater relative contribution from actor–critic-type learning. Importantly, these results conceptually replicate, for the first time, our previous work, in which we showed a decreased contribution of Q-learning during a gain-seeking/loss-avoidance task (6). In the current study, performance impairments were also in part related to increased undirected noise, which accounts for nondeterministic choices even in the face of strong evidence. We have observed this in previous RL studies (15), and in the current study it was mostly associated with interindividual differences in IQ.

Which mechanisms could underlie a selective impairment in the representation of expected value? Decreased learning from gains, as opposed to intact loss avoidance, has been identified as one potential mechanism (6,14,29,30). In this study, impaired performance on more deterministic pairs, associated with more gains than neutral outcomes, but spared performance on 60-40 trials, where learning occurs almost equally from gains and neutral outcomes, provides circumstantial evidence for this notion. One improvement compared with

Figure 2. Learning and transfer phase performance. **(A–E)** Solid bars represent healthy volunteers (HV); bars with diagonal lines represent people with schizophrenia (PSZ). *$p < .05$, **$p < .01$, ***$p < .001$, ^{a}trend ($p = .06$–$.09$). Error bars represent 95% confidence interval except for learning phase data, where bars represent SEM. Asterisks above error bars represent significant preference against chance; asterisks above solid horizontal lines represent between-group or within-group differences. In panel **(E)** (center), 75D vs. All/75 vs. All shows performance on 75D/P trials vs. all choice options of nonidentical value. In panel **(E)** (right), separate plots for 75D and 75P versus other choice options broken down by their value difference (x-axis) are shown.

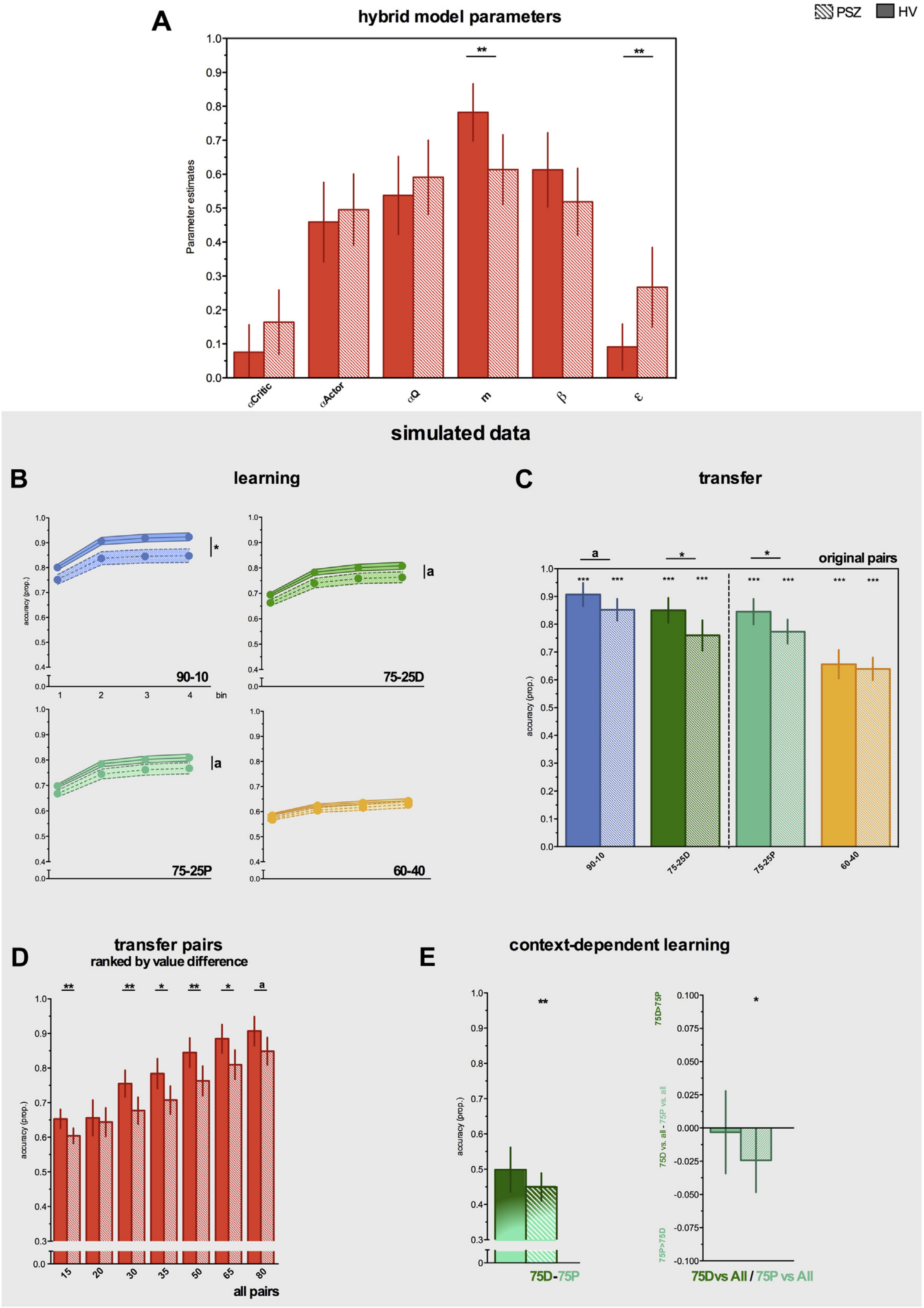


Figure 3. **(A–E)** Hybrid model parameters and simulated data: n(simulations) = 250, simulated with undirected noise (ε) of 50%. β, inverse temperature; HV, healthy volunteers; m, mixing; PSZ, people with schizophrenia. *$p < .05$, **$p < .01$, ***$p < .001$, ^{a}trend (p = .06–.09).

Table 2. Hybrid Model Parameters per Participant Group

	HVs (*n* = 36)	PSZ (*n* = 42)	*t*	*p*
Critic Learning Rate (α_c)	.08 (.24)	.16 (.31)	1.40	.17
Actor Learning Rate (α_a)	.46 (.35)	.50 (.34)	0.47	.64
Q-Learning Rate (α_q)	.54 (.34)	.59 (.34)	0.69	.50
Mixing Parameter (*m*)	.78 (.25)	.61 (.33)	2.51	.01
Inverse Temperature (β)	.61 (.32)	.52 (.32)	1.30	.20
Undirected Noise (ε)	.09 (.20)	.27 (.38)	2.52	.01

Values are presented as mean (SD).

HVs, healthy volunteers; PSZ, people with schizophrenia.

previous paradigms is that here we focused on reward value instead of contrasting valence conditions, which is a direct test of expected value deficits. The current results show, for the first time, that a diminished role of expected value in driving choices can lead to suboptimal behavior in a dose-response fashion; that is, performance impairments increase monotonically with increased demands placed on expected value computations. This work further strengthens the claim that deficits in the representation of expected value are a central feature of learning and decision-making impairments in PSZ, and here we reveal when these deficits should be most evident.

The relationship between the value difference effect and Q-learning fits well with previous neuroimaging studies. Work from our group has identified attenuated expected value signals in insula and anterior cingulate, regions that encode (state-dependent) expected value (31,32), in PSZ with motivational deficits (5,14). Ventromedial and orbitofrontal prefrontal cortex dysfunction, consistently involved in tracking reward value (8,9,33), has also been linked to learning and decision-making deficits in schizophrenia (34,35). Thus, a diminished role for expected value in decision making, demonstrated by the value difference effect and confirmed by our computational model, is suggestive of impairments in a range of cortical areas that encode reward value.

We have argued that underuse of expected value and a relative increase in reliance on stimulus-response learning can also enhance the effect of context on stimulus valuation, leading to a unique prediction in which preferences can arise among choice options with identical reinforcement probabilities. For this hypothesis, we found subtle but consistent evidence only in PSZ, but not in HVs, which could selectively be accounted for by a context-dependent state-value RPE (via α_c). Interestingly, the magnitude of this context parameter was greater in individuals with high motivational deficits. Given that there was no association between motivational deficits and the mixing parameter (or between the mixing and context parameters), this result implies that a context-dependent choice bias in PSZ with motivational deficits can be observed only to the degree that they rely on actor–critic-type learning. This may suggest that increased sensitivity to context and impairments in Q-learning may be differentially sensitive to symptom severity and patient status, respectively.

Although the effect of contextual reward availability on decision making was subtle in PSZ, these findings are noteworthy. Klein *et al.* (36) revealed that learning the value of one stimulus relative to another can lead to suboptimal decision making. In their study, a relative RPE signal was specifically encoded by the striatum. Despite clear differences between the task design of Klein *et al.* and the current study, most notably pairwise versus blockwise context effects, their work does provide evidence for the notion that the effect of context on perceived stimulus value seems to be encoded specifically by brain regions typically associated with RPE signaling.

Related to this point, we observed intact learning on 60-40 trials in PSZ [see also Waltz *et al.* (28)], which improved gradually and relies on slow accumulation of RPEs (18). Subtle evidence for a context effect, a relative increase in the contribution of actor–critic-type learning, and adequate learning on 60-40 trials are consistent with relatively intact striatal function in our medicated sample. These findings align well with intact striatal RPE signaling in medicated PSZ (37) as well as normalization of reward signals following treatment with antipsychotics (38).

It is interesting to speculate on how impairments in stimulus-response learning and expected value may change with illness phase or medication status. In nonmedicated and/or first-episode patients, abnormal RPE signals in striatum and

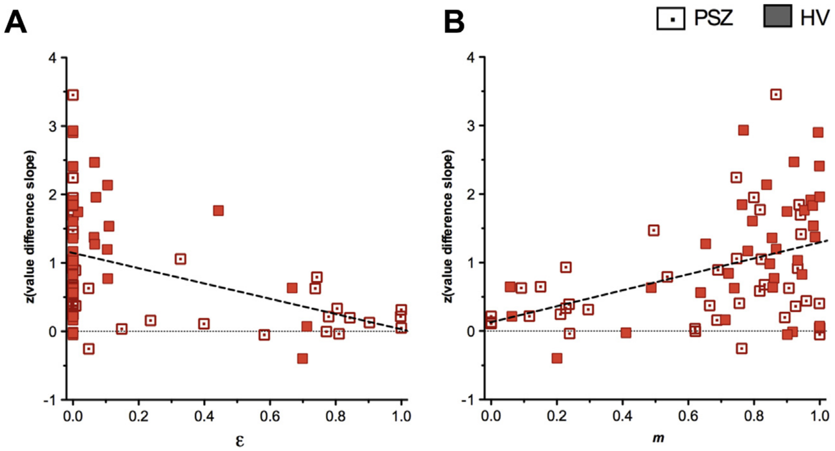


Figure 4. **(A, B)** Significant correlations between model parameters and task performance. ε, undirected noise; HV, healthy volunteers; *m*, mixing; PSZ, people with schizophrenia.

midbrain have been reported (4,39,40), while a recent study did not observe differences in striatal RPE signals between HVs and PSZ on long-term medication (37). In addition, there exists some evidence that deficits in expected value can be observed in both first-episode patients (16) and long-term PSZ (6,29,41). In the absence of correlations between antipsychotic dose and Q-learning, this work may suggest that deficits in the representation of expected value exist across the psychosis spectrum, while impaired stimulus-response learning may be especially pronounced in the early phase of the illness and perhaps rescued by antipsychotic medication. While disentangling illness phase from medication effects is an arduous task, such studies may ultimately provide much-needed insights into symptom mechanisms across the psychosis spectrum.

To summarize, this work provides specific evidence that decision-making impairments in PSZ increase monotonically with demands placed on expected value computations. A greater influence of stimulus-response learning as a result of underuse of expected value may produce additional violations of optimal decision-making policies such as a contextual or relative choice bias. This work provides a novel source of evidence suggesting a diminished role of expected value in guiding optimal decisions in PSZ and sheds light on the conditions that facilitate such impairments.

Limitations

Some limitations warrant discussion. While we were able to replicate our previous finding of decreased Q-learning or relative increase in actor–critic-type learning in PSZ (6), the mixing parameter was not associated with symptom ratings. Previous studies investigating RL deficits in PSZ have reported mixed results regarding relationships to negative symptoms (6,16,29). Compared with our previous study (6), here we used a wide range of choice pairs and a comprehensive transfer phase. Greater demands placed on expected value computations may have increased sensitivity to detect group differences, as opposed to differences in HVs and PSZ with high motivational deficits only. Moreover, the use of a context-dependent learning rate for the critic, which was associated with motivational deficits, may have explained some of the variance that would have otherwise been captured by other model parameters. While multiple factors may explain the absence of an association between the mixing parameter and motivational deficit severity, the current study results still provides evidence for the notion that expected value deficits are an essential part of schizophrenia.

It should also be noted that an alternative account of the current findings is that PSZ may rely less on model-based strategies (42). Both Q-based and model-based learning make identical predictions for this task; that is, Q-learning predicts improved performance at greater levels of value difference via action-value learning, while model-based strategies predict improved performance when action-outcome sequences are better understood. Importantly, this alternative explanation does not change the interpretation of increased reliance on model-free stimulus-response learning in PSZ.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the National Institute of Mental Health (Grant No. MH80066 to JMG).

We thank Benjamin M. Robinson for his contributions to the task design.

JAW, JMG, and MJF perform consulting for Hoffman-La Roche. JMG has also consulted for Takeda and Lundbeck and receives royalty payments from the Brief Assessment of Cognition in Schizophrenia. JAW also consults for NCT Holdings. The current experiments were not related to any consulting activity. DH reports no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Maryland Psychiatric Research Center (DH, JMG, JAW), Department of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland, and Department of Cognitive, Linguistic & Psychological Sciences and Department of Psychiatry and Human Behavior (MJF), Brown University, Providence, Rhode Island.

Address correspondence to Dennis Hernaus, Ph.D., Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, P.O. Box 21247, Baltimore, MD 21228; E-mail: dhernaus@som.umaryland.edu.

Received Dec 19, 2017; revised and accepted Mar 20, 2018.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.bpsc.2018.03.014.

REFERENCES

1. Heinz A, Schlagenhauf F (2010): Dopaminergic dysfunction in schizophrenia: Salience attribution revisited. Schizophr Bull 36:472–485.
2. Waltz JA, Gold JM (2016): Motivational deficits in schizophrenia and the representation of expected value. Curr Top Behav Neurosci 27:375–410.
3. Morris RW, Vercammen A, Lenroot R, Moore L, Langton JM, Short B, *et al.* (2012): Disambiguating ventral striatum fMRI-related BOLD signal during reward prediction in schizophrenia. Mol Psychiatry 17:235. 280–289.
4. Murray GK, Corlett PR, Clark L, Pessiglione M, Blackwell AD, Honey G, *et al.* (2008): Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol Psychiatry 13:239. 267–276.
5. Waltz JA, Schweitzer JB, Gold JM, Kurup PK, Ross TJ, Salmeron BJ, *et al.* (2009): Patients with schizophrenia have a reduced neural response to both unpredictable and predictable primary reinforcers. Neuropsychopharmacology 34:1567–1577.
6. Gold JM, Waltz JA, Matveeva TM, Kasanova Z, Strauss GP, Herbener ES, *et al.* (2012): Negative symptoms and the failure to represent the expected reward value of actions: Behavioral and computational modeling evidence. Arch Gen Psychiatry 69:129–138.
7. Barch DM, Treadway MT, Schoen N (2014): Effort, anhedonia, and function in schizophrenia: Reduced effort allocation predicts amotivation and functional impairment. J Abnorm Psychol 123:387–397.
8. Padoa-Schioppa C, Cai X (2011): The orbitofrontal cortex and the computation of subjective value: Consolidated concepts and new perspectives. Ann N Y Acad Sci 1239:130–137.
9. Clithero JA, Rangel A (2014): Informatic parcellation of the network involved in the computation of subjective value. Soc Cogn Affect Neurosci 9:1289–1302.
10. Hogeveen J, Hauner KK, Chau A, Krueger F, Grafman J (2017): Impaired valuation leads to increased apathy following ventromedial prefrontal cortex damage. Cereb Cortex 27:1401–1408.
11. Schultz W, Dayan P, Montague PR (1997): A neural substrate of prediction and reward. Science 275:1593–1599.
12. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013): A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966–973.
13. Dowd EC, Frank MJ, Collins A, Gold JM, Barch DM (2016): Probabilistic reinforcement learning in patients with schizophrenia: Relationships to anhedonia and avolition. Biol Psychiatry Cogn Neurosci Neuroimaging 1:460–473.

14. Waltz JA, Xu Z, Brown EC, Ruiz RR, Frank MJ, Gold J (2018): Motivational deficits in schizophrenia are associated with reduced differentiation between gain and loss-avoidance feedback in the striatum. Biol Psychiatry Cogn Neurosci Neuroimaging 3:238–247.
15. Collins AG, Brown JK, Gold JM, Waltz JA, Frank MJ (2014): Working memory contributions to reinforcement learning impairments in schizophrenia. J Neurosci 34:13747–13756.
16. Chang WC, Waltz JA, Gold JM, Chan TCW, Chen EYH (2016): Mild reinforcement learning deficits in patients with first-episode psychosis. Schizophr Bull 42:1476–1485.
17. Watkins C, Dayan P (1992): Q-learning. Mach Learn 8:279–292.
18. Frank MJ, Claus ED (2006): Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol Rev 113:300–326.
19. Roesch MR, Olson CR (2007): Neuronal activity related to anticipated reward in frontal cortex: Does it represent value or reflect motivation? Ann N Y Acad Sci 1121:431–446.
20. Sutton RS, Barto AG (1998): Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press.
21. Joel D, Niv Y, Ruppin E (2002): Actor-critic models of the basal ganglia: New anatomical and computational perspectives. Neural Netw 15:535–547.
22. Collins AG, Frank MJ (2014): Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol Rev 121:337–366.
23. Calabresi P, Gubellini P, Centonze D, Picconi B, Bernardi G, Chergui K, *et al.* (2000): Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. J Neurosci 20:8443–8451.
24. First MB, Spitzer RL, Gibbon M, Williams JBW (1997): Structured Clinical Interview for DSM-IV-Axis I Disorders (SCID-I). Washington, DC: American Psychiatric Press.
25. Pfohl B, Blum N, Zimmerman M, Stangl D (1989): Structured Interview for DSM-III-R Personality Disorders (SIDP-R). Iowa City, IA: University of Iowa, Department of Psychiatry.
26. Andreasen NC (1984): The Scale for the Assessment of Negative Symptoms (SANS). Iowa City, IA: University of Iowa.
27. McMahon RP, Kelly DL, Kreyenbuhl J, Kirkpatrick B, Love RC, Conley RR (2002): Novel factor-based symptom scores in treatment resistant schizophrenia: Implications for clinical trials. Neuropsychopharmacology 26:537–545.
28. Waltz JA, Frank MJ, Robinson BM, Gold JM (2007): Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. Biol Psychiatry 62:756–764.
29. Hartmann-Riemer MN, Aschenbrenner S, Bossert M, Westermann C, Seifritz E, Tobler PN, *et al.* (2017): Deficits in reinforcement learning but no link to apathy in patients with schizophrenia. Sci Rep 7:40352.
30. Maia TV, Frank MJ (2017): An integrative perspective on the role of dopamine in schizophrenia. Biol Psychiatry 81:52–66.
31. Becker CA, Flaisch T, Renner B, Schupp HT (2017): From thirst to satiety: The anterior mid-cingulate cortex and right posterior insula indicate dynamic changes in incentive value. Front Hum Neurosci 11:234.
32. Rolls ET, McCabe C, Redoute J (2008): Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. Cereb Cortex 18:652–663.
33. Metereau E, Dreher JC (2015): The medial orbitofrontal cortex encodes a general unsigned value signal during anticipation of both appetitive and aversive events. Cortex 63:42–54.
34. Schlagenhauf F, Sterzer P, Schmack K, Ballmaier M, Rapp M, Wrase J, *et al.* (2009): Reward feedback alterations in unmedicated schizophrenia patients: Relevance for delusions. Biol Psychiatry 65:1032–1039.
35. Park IH, Lee BC, Kim JJ, Kim JI, Koo MS (2017): Effort-based reinforcement processing and functional connectivity underlying amotivation in medicated patients with depression and schizophrenia. J Neurosci 37:4370–4380.
36. Klein TA, Ullsperger M, Jocham G (2017): Learning relative values in the striatum induces violations of normative decision making. Nat Commun 8:16033.
37. Culbreth AJ, Westbrook A, Xu Z, Barch DM, Waltz JA (2016): Intact ventral striatal prediction error signaling in medicated schizophrenia patients. Biol Psychiatry Cogn Neurosci Neuroimaging 1:474–483.
38. Nielsen MO, Rostrup E, Wulff S, Bak N, Broberg BV, Lublin H, *et al.* (2012): Improvement of brain reward abnormalities by antipsychotic monotherapy in schizophrenia. Arch Gen Psychiatry 69:1195–1204.
39. Reinen JM, Van Snellenberg JX, Horga G, Abi-Dargham A, Daw ND, Shohamy D (2016): Motivational context modulates prediction error response in schizophrenia. Schizophr Bull 42:1467–1475.
40. Schlagenhauf F, Huys QJ, Deserno L, Rapp MA, Beck A, Heinze HJ, *et al.* (2014): Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage 89:171–180.
41. Brown EC, Hack SM, Gold JM, Carpenter WT Jr, Fischer BA, Prentice KP, *et al.* (2015): Integrating frequency and magnitude information in decision-making in schizophrenia: An account of patient performance on the Iowa Gambling Task. J Psychiatr Res 66–67:16–23.
42. Culbreth AJ, Westbrook A, Daw ND, Botvinick M, Barch DM (2016): Reduced model-based decision-making in schizophrenia. J Abnorm Psychol 125:777–787.