Negative Symptoms and the Failure to Represent the Expected Reward Value of Actions

Behavioral and Computational Modeling Evidence

James M. Gold, PhD; James A. Waltz, PhD; Tatyana M. Matveeva, BA; Zuzana Kasanova, BA; Gregory P. Strauss, PhD; Ellen S. Herbener, PhD; Anne G. E. Collins, PhD; Michael J. Frank, PhD

Context: Negative symptoms are a core feature of schizophrenia, but their pathogenesis remains unclear. Negative symptoms are defined by the absence of normal function. However, there must be a productive mechanism that leads to this absence.

Objective: To test a reinforcement learning account suggesting that negative symptoms result from a failure in the representation of the expected value of rewards coupled with preserved loss-avoidance learning.

Design: Participants performed a probabilistic reinforcement learning paradigm involving stimulus pairs in which choices resulted in reward or in loss avoidance. Following training, participants indicated their valuation of the stimuli in a transfer test phase. Computational modeling was used to distinguish between alternative accounts of the data.

Setting: A tertiary care research outpatient clinic.

Patients: In total, 47 clinically stable patients with a diagnosis of schizophrenia or schizoaffective disorder and 28 healthy volunteers participated in the study. Patients were divided into a high-negative symptom group and a low-negative symptom group.

Main Outcome Measures: The number of choices leading to reward or loss avoidance, as well as performance in the transfer test phase. Quantitative fits from 3 different models were examined.

Results: Patients in the high-negative symptom group demonstrated impaired learning from rewards but intact loss-avoidance learning and failed to distinguish rewarding stimuli from loss-avoiding stimuli in the transfer test phase. Model fits revealed that patients in the high-negative symptom group were better characterized by an "actor-critic" model, learning stimulus-response associations, whereas control subjects and patients in the low-negative symptom group incorporated expected value of their actions ("Q learning") into the selection process.

Conclusions: Negative symptoms in schizophrenia are associated with a specific reinforcement learning abnormality: patients with high-negative symptoms do not represent the expected value of rewards when making decisions but learn to avoid punishments through the use of prediction errors. This computational framework offers the potential to understand negative symptoms at a mechanistic level.

Arch Gen Psychiatry. 2012;69(2):129-138

IN THE PAST DECADE, INTEREST IN the role of deficits in reinforcement learning (RL) and reward processing for understanding the symptoms of schizophrenia has been increasing.1-4 This work has been shaped by studies5-6 of behaving primates showing that the pattern of dopamine cell firing seems to code reward prediction errors (PEs), with cells increasing their phasic firing rates when outcomes are better than expected (positive PEs) and briefly ceasing to fire when outcomes are worse than expected (negative PEs). It is thought that positive PE signals are broadcast to dopamine cell target areas and serve to reinforce currently active motor responses and representations that are associated with better-than-expected outcomes. In contrast, transient cessations in dopamine cell activity indicate that current actions have resulted in poorer-than-expected outcomes and should be avoided. This pattern of dopamine cell firing has been successfully modeled using RL algorithms,7-9 and there is consistent support for the notion that phasic dopamine signals modify synaptic plasticity in corticostriatal circuits associated with action selection.10-12

It is well documented that antipsychotic medications achieve their effect through the blockade of dopamine receptors, supporting the inference that psychosis is linked to excessive dopamine release.13-16 It has been proposed that excessive dopamine cell firing might “reinforce” or inappropriately increase the sa-
of stimuli and responses, driving aberrant learning processes that contribute to psychosis.\(^3,^{37}\) This hypothesis has been supported by empirical evidence indicating associations between abnormal PE signaling and the presence of psychosis\(^18\) and the severity of delusions.\(^9\)

Waltz et al\(^{20}\) argued that the negative symptoms of schizophrenia may be understood as reflecting a different RL abnormality.\(^7\) It was found that patients with schizophrenia show reduced learning from positive outcomes compared with control subjects but do not differ from control subjects in learning from negative outcomes.\(^{10}\) This pattern of performance is most pronounced in patients with high-negative symptoms. However, it is unclear whether impairments in learning from positive outcomes reflect impaired learning from positive PEs per se or a deficit in representing the expected reward value of choices themselves. Past investigations used learning tasks in which the optimal response was associated with reward receipt that should generate a positive PE and the less optimal choice was associated with an actual loss or withholding an expected reward, both of which should generate negative PEs.\(^{20,22}\) Therefore, these earlier studies could not distinguish between a failure to learn from positive PEs and a failure in the representation of the prospective reward values during decision making.

This distinction between positive PEs and the valuation of positive outcomes when making choices maps onto the distributed neural system that is involved with decision making. In the basal ganglia, reinforcement outcomes influence subsequent behavioral choices through synaptic plasticity in response to PEs signaled by dopamine neurons. This “slow” learning system is complemented by the contribution of the orbitofrontal cortex (OFC),\(^{21}\) thought to represent the expected value of potential outcomes in working memory.\(^{22,23}\) These OFC value representations are believed to be more rapidly and flexibly updated than those in the basal ganglia and provide additional “top-down” influence on decision making. Therefore, reward-based decision making involves (at least) 2 separate processes, namely, a learning mechanism that reinforces choices that have led to positive PEs in the past and a representation of the expected value of a situation-action pair.\(^{23-25}\)

In computational models of reward learning and decision making, the contribution of the basal ganglia system is often formalized using an “actor-critic” framework or a “Q learning” framework.\(^9\) In the former, a “critic” evaluates the reward values of particular states, and the “actor” selects responses as a function of learned stimulus-response weights. When outcomes differ from expectations, PEs are used to modify learning in the critic itself (to better predict reward values in the future). The critic’s PEs also serve to increase and decrease stimulus-response weights in the actor. With learning, this scheme allows the actor to select actions with strong stimulus-response weights (ie, those that have produced more positive than negative PEs) without representing the expected reward values of the actions themselves. In contrast, in Q learning, instead of learning the value of particular states, the expected quality (“Q value”) of each action is learned separately; actions are selected by comparing the various Q values of each candidate action and probabilistically choosing the largest one. In this case, PEs are computed with respect to the expected Q value of the selected action and are used to adjust expected action value directly. Therefore, whereas the actor in the actor-critic scheme does not consider the outcome values of competing actions, the Q-learning scheme makes these fundamental. There is compelling evidence that the OFC has a critical role in representing these kinds of value representations.\(^{23,26-28}\)

Although these 2 RL algorithms are similar,\(^9\) they make different predictions about the nature of representations used in reward-based decision making that can be highlighted with appropriate task manipulations. Herein, we examine a hybrid model in which (putatively striatal) action weights are learned as a function of PEs but are also modulated by representation of the expected outcome Q value (due to putative top-down input from the OFC).\(^{23}\)

Moreover, this modeling framework provides an important means of contrasting different hypotheses about the origins of the reward learning deficits found in earlier work. Specifically, deficits in learning from rewarding outcomes could be the result of a primary failure in the ability to signal positive PEs or, alternatively, impairments in the ability to represent the positive expected value of decision outcomes.

Following work by Pessiglione et al\(^{29}\) and Kim et al,\(^{30}\) we implemented a task in which participants were asked to simultaneously learn 4 discriminations. In 2 pairs, the choice of the optimal stimulus is probabilistically associated with the receipt of money (a positive PE), and the choice of the nonoptimal stimulus results in no reward (ie, a zero outcome). Failure to obtain a reward in these pairs should result in a negative PE. In these pairs, as in prior work, rewards and positive PEs are conflated. In 2 other pairs, the choice of the optimal stimulus results in no loss (ie, loss avoidance, a zero outcome), whereas the choice of the nonoptimal stimulus results in overt monetary loss. In this overall design, the same “zero” outcome would result in a positive PE when it occurs in the context of potential negative outcomes\(^{31,32}\) but would result in a negative PE when encountered in the context of potential rewarding outcomes. This interpretation is consistent with computational models indicating that active avoidance relies on learning from positive PEs\(^{33-35}\) and that avoidance of an aversive outcome activates reward areas.\(^{30}\)

After the initial acquisition phase of the task, participants completed a transfer test phase in which they chose between novel combinations of all the trained stimuli without additional feedback.\(^{36}\) Critically, some pairs involved selecting between an action that had been rewarding and one that had simply avoided a loss. Both actions should have produced positive PEs during learning, leading to an increased tendency to select them. If one’s choices are solely determined by the strength of association with positive PEs (as in actor-critic), the rewarding stimulus and loss-avoiding stimulus would be of equal value. Alternatively, if one was sensitive to the expected outcome of the action (eg, if action selection relies on Q values), participants should prefer the gain-producing action over the action with zero outcome.

We hypothesized that patients having schizophrenia with high-negative symptoms would show specific impairment in representing reward value. This would likely implicate OFC dysfunction in these patients.\(^{20,37}\)
Participating in the study. All participants were divided into a high-negative symptom (HNS) group and a low-negative symptom group (LNS) by a median split on the sum of the avolition and anhedonia global items on the Scale for the Assessment of Negative Symptoms. These items were selected because they have been found to reflect a single factor and are more theoretically relevant to reward items than the “restricted affect” factor of the Scale for the Assessment of Negative Symptoms.

The 3 groups did not significantly differ in age, sex, or race/ethnicity (Table). The HC group had more years of education than both patient groups, and the LNS group completed more years of education than the HNS group. Both patient groups had moderate symptom severity, as indicated by their Brief Psychiatric Rating Scale total scores. When Brief Psychiatric Rating Scale factors were examined, the HNS group had greater severity on the negative cluster symptom factor, while the 2 patient groups did not differ on the positive cluster or disorganized cluster factors.

Neuropsychological Testing

All participants completed measures of word reading, general intelligence, and the MATRICS (Measurement and Treatment Research to Improve Cognition in Schizophrenia Consensus Cognitive) battery. The HC group scored significantly higher than both patient groups on all word reading measures, including the Wide Range Achievement Test (WRAT), the Wechsler Test of Adult Reading (WTAR), and the Abbreviated Scale of Intelligence (WASI). The HC group also scored higher than both patient groups on both the positive and negative clusters of the Brief Psychiatric Rating Scale (BPRS). Neither the HNS group nor the LNS group differed significantly from the HC group on these measures.

The participants in the HNS group had lower scores on the MATRICS battery than the HC group, while the LNS group had similar scores to the HC group. The MATRICS battery measures several aspects of cognitive function, including memory, attention, learning, and executive function. The lower scores in the HNS group suggest that they may have more cognitive impairments than the HC group.

In conclusion, the study found that patients with schizophrenia have significant cognitive impairments, especially in the negative clusters of the BPRS and MATRICS battery. The HNS group had more severe symptoms than the LNS group, and both patient groups had lower cognitive scores than the HC group. These findings suggest that cognitive impairment is a common feature of schizophrenia, and that it may be more severe in patients with more severe symptoms.
higher than both patient groups on all standard measures (Table). The 2 patient groups showed almost identical performance on all standard cognitive measures.

**RL TASK**

The learning task was administered via commercially available software (E-Prime; Psychology Software Tools) and was run on a laptop computer with a 17-in monitor. Stimuli were color images of landscapes appearing on a gray background. Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Following training, the transfer test phase was presented. In these 64 trials, the original 4 training pairs were each presented 4 times, and the 24 novel pairings were each presented twice. For novel pairings, each trained item was presented with every other trained item (ie, an item that had been a 90% winner was paired with both items from the 80% gain pair, the 90% loss-avoidance pair, and the 80% loss-avoidance pair). Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!,” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Following training, the transfer test phase was presented. In these 64 trials, the original 4 training pairs were each presented 4 times, and the 24 novel pairings were each presented twice. For novel pairings, each trained item was presented with every other trained item (ie, an item that had been a 90% winner was paired with both items from the 80% gain pair, the 90% loss-avoidance pair, and the 80% loss-avoidance pair). Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!,” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Following training, the transfer test phase was presented. In these 64 trials, the original 4 training pairs were each presented 4 times, and the 24 novel pairings were each presented twice. For novel pairings, each trained item was presented with every other trained item (ie, an item that had been a 90% winner was paired with both items from the 80% gain pair, the 90% loss-avoidance pair, and the 80% loss-avoidance pair). Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!,” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Following training, the transfer test phase was presented. In these 64 trials, the original 4 training pairs were each presented 4 times, and the 24 novel pairings were each presented twice. For novel pairings, each trained item was presented with every other trained item (ie, an item that had been a 90% winner was paired with both items from the 80% gain pair, the 90% loss-avoidance pair, and the 80% loss-avoidance pair). Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!,” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Following training, the transfer test phase was presented. In these 64 trials, the original 4 training pairs were each presented 4 times, and the 24 novel pairings were each presented twice. For novel pairings, each trained item was presented with every other trained item (ie, an item that had been a 90% winner was paired with both items from the 80% gain pair, the 90% loss-avoidance pair, and the 80% loss-avoidance pair). Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!,” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Following training, the transfer test phase was presented. In these 64 trials, the original 4 training pairs were each presented 4 times, and the 24 novel pairings were each presented twice. For novel pairings, each trained item was presented with every other trained item (ie, an item that had been a 90% winner was paired with both items from the 80% gain pair, the 90% loss-avoidance pair, and the 80% loss-avoidance pair). Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!,” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Following training, the transfer test phase was presented. In these 64 trials, the original 4 training pairs were each presented 4 times, and the 24 novel pairings were each presented twice. For novel pairings, each trained item was presented with every other trained item (ie, an item that had been a 90% winner was paired with both items from the 80% gain pair, the 90% loss-avoidance pair, and the 80% loss-avoidance pair). Participants were presented with 4 pairs of landscape items, 1 pair at a time (Figure 1). Two pairs involved potential gain; if the correct item was selected, participants saw an image of a nickel coupled with the word “Win!,” whereas if the incorrect item was selected, they saw “Not a winner, Try again!” The correct response was reinforced on 90% of trials in one pair and on 80% of trials in the other pair. Two other pairs involved learning to avoid losses; in these pairs, selection of the correct response received the feedback “Keep your money!,” whereas selection of the incorrect item resulted in the feedback “Lose!” Therefore, if the best item in the loss-avoiding pairs was selected, participants avoided a loss 90% or 80% of the time. A brief 12-trial practice session was administered to ensure task comprehension, followed by 160 learning trials with all pair types presented in a randomized order. Each pair was shown 40 times during training. To examine learning, the 160 trials were divided into 4 learning blocks of 40 trials.

Figure 1. Example of reinforcement learning task stimuli and feedback. A, Feedback delivered after a correct choice (indicated by a blue border) in the reward trials. B, Feedback delivered following an incorrect choice. C, Feedback delivered following a correct choice in the loss-avoidance trials. D, Feedback delivered following an incorrect choice.
pants were instructed to pick the item in the pair that they thought was “best” based on their earlier learning. No feedback was administered during this phase.

**COMPUTATIONAL MODEL**

We examined the ability of the following 3 different models to fit each participant’s trial-by-trial sequence of choices across training and transfer test phases: (1) a standard actor-critic architecture simulating pure basal ganglia–dependent learning, (2) a pure Q-learning model simulating action selection as a function of learned expected reward value, and (3) a hybrid model in which an actor-critic is “augmented” by a Q-learning component meant to capture the top-down influence of OFC value representations onto striatum. See the Appendix, eFigure 1, and eFigure 2 (http://www.archgenpsychiatry.com).

**STATISTICAL ANALYSIS**

An omnibus repeated-measures analysis of variance (ANOVA) was first conducted with a between-subject factor of group (HC, LNS, and HNS) and within-subject factors for feedback valence (gain vs loss avoidance), probability (90% and 80%), and learning block (blocks 1-4). Hyun-Feldt correction was applied if assumption of sphericity was violated; unless indicated, sphericity was not violated, and no correction was made. Significant interactions were followed by a series of ANOVA and post hoc least significant difference (LSD) test contrasts examining differences in block 4 performances. To examine the balance of learning from gain vs loss avoidance, we subtracted block 4 learning achieved from gain from that achieved from loss avoidance, testing for group differences using ANOVA, followed by within-group paired-sample t test. Transfer test phase performance was examined using 1-way ANOVA, followed by LSD post hoc contrasts.

**RESULTS**

**BEHAVIORAL FINDINGS**

As shown in **Figure 2A**, the HC group and the LNS group demonstrated robust learning in the 90% gain condition, with the HNS group demonstrating limited learning. In contrast, the groups performed similarly in the 80% gain condition (Figure 2B). The loss-avoidance learning blocks produced different results. Here, the HNS group matched or performed at slightly higher levels than the other 2 groups, suggesting that their learning is more effectively driven by loss avoidance than by gain seeking (Figure 2C and D).

![Figure 2](http://www.archgenpsychiatry.com)

**Figure 2.** Differences in reinforcement learning among patients and healthy control (HC) subjects in 90% and 80% probability gain and loss-avoidance conditions. A and B, Performance in the 90% and 80% gain conditions, respectively. C and D, Performance in the 90% and 80% loss-avoidance conditions, respectively. HNS indicates high-negative symptom; LNS, low-negative symptom.
was a significant feedback valence–X–learning block interaction ($F_{6,72}=4.42$, $P = .005$), qualified by a trend toward a feedback valence–X–learning block–X–group interaction ($F_{6,72}=2.22$, $P = .06$ after Hyun-Feldt correction). This last interaction suggests that the groups learned differently over time as a function of whether they were learning from rewards or from loss avoidance.

To assess whether feedback valence differentially affected final performance levels, we conducted a 2 feedback valence–X–2 probability–X–3 group repeated-measures ANOVA with block 4 performance as the dependent variable because it captured asymptotic learning levels. This analysis produced a significant main effect of probability ($F_{1,72}=4.77$, $P = .03$ [90% greater than 80% stimuli]) and a significant group–X–feedback valence interaction ($F_{2,72}=4.51$, $P = .01$) (ie, the groups learned differently as a function of feedback valence). The probability–X–group interaction fell short of significance ($F_{2,72}=2.43$, $P = .10$); no other effects approached significance. One-way ANOVAs examining performance for each of the 4 stimulus pairs were conducted to explore the nature of the differences observed among the 3 groups (overall differences were significant ($F_{2,72}=5.81$, $P = .005$), with post hoc LSD comparisons indicating a significant difference between the HC group and the HNS group ($P = .001$) and an approach toward a significant difference between the LNS group and the HNS group ($P = .06$). As shown in Figure 4C, the HC group showed a robust preference for frequently rewarded stimuli over loss avoiders, consistent with the pattern expected if they were representing the positive expected value of the stimuli rather than relying on the number of times a stimulus has been associated with a positive PE. In contrast, the HNS group showed no preference for gain relative to loss avoiders, indicating that their preferences were based on the accumulation of positive PEs and did not take into account the value associated with those positive PEs. Although we assessed whether there were significant differences between groups in other stimulus–feedback valence comparisons, no other statistically significant differences were found.

An alternative explanation for our results is that the lack of preference for gain over loss avoidance in the HNS group might be due to difficulty in learning about rewards in general. However, as shown in Figure 4, the HNS group demonstrated a robust preference for frequently rewarded stimuli over frequently losing (FW vs FL in the figure) stimuli during the transfer test phase, with no differences observed among the 3 groups (overall $F_{2,72}=2.06$, $P = .14$). Furthermore, the HNS group preferred frequently rewarded stimuli over infrequently rewarded stimuli (FW vs IW in the figure). Therefore, the failure to prefer “winners” over loss avoiders cannot be explained by a failure to have learned which stimuli were associated with reward receipt.

We also examined the preference for frequent loss avoiders over infrequent winners (FLA vs IW in the figures). All 3 groups had a robust preference for the loss avoiders, despite the fact that the infrequent winner actually had a slightly positive expected value that was higher than that of loss avoiders. Therefore, all 3 groups preferred the stimulus that was more frequently associated with a positive PE over a choice that had a higher expected value but was also associated with more frequent negative PEs during learning.

**TRANSFER TEST PHASE PERFORMANCE**

Performance on 9 types of novel stimulus pairings was examined for the transfer test phase (Appendix, eFigure 1, eFigure 2, and Figure 4C). Pairings in which participants were confronted with the most frequently rewarded stimuli (FW in the figures) and the stimuli that most reliably avoided losses (FLA in the figures) provided the critical test of the hypothesis that the HNS group showed a specific impairment in representation of expected positive value of decision outcomes rather than learning from positive PEs. The 1-way ANOVA examining differences among the groups was significant ($F_{2,72}=5.81$, $P = .005$), with post hoc LSD comparisons indicating a significant difference between the HC group and the HNS group ($P = .001$) and an approach toward a significant difference between the LNS group and the HNS group ($P = .06$). As shown in Figure 4C, the HC group showed a robust preference for frequently rewarded stimuli over loss avoiders, consistent with the pattern expected if they were representing the positive expected value of the stimuli rather than relying on the number of times a stimulus has been associated with a positive PE. In contrast, the HNS group showed no preference for gain relative to loss avoiders, indicating that their preferences were based on the accumulation of positive PEs and did not take into account the value associated with those positive PEs. Although we assessed whether there were significant differences between groups in other stimulus–feedback valence comparisons, no other statistically significant differences were found.

An alternative explanation for our results is that the lack of preference for gain over loss avoidance in the HNS group might be due to difficulty in learning about rewards in general. However, as shown in Figure 4, the HNS group demonstrated a robust preference for frequently rewarded stimuli over frequently losing (FW vs FL in the figure) stimuli during the transfer test phase, with no differences observed among the 3 groups (overall $F_{2,72}=2.06$, $P = .14$). Furthermore, the HNS group preferred frequently rewarded stimuli over infrequently rewarded stimuli (FW vs IW in the figure). Therefore, the failure to prefer “winners” over loss avoiders cannot be explained by a failure to have learned which stimuli were associated with reward receipt.

We also examined the preference for frequent loss avoiders over infrequent winners (FLA vs IW in the figures). All 3 groups had a robust preference for the loss avoiders, despite the fact that the infrequent winner actually had a slightly positive expected value that was higher than that of loss avoiders. Therefore, all 3 groups preferred the stimulus that was more frequently associated with a positive PE over a choice that had a higher expected value but was also associated with more frequent negative PEs during learning.
EFFECTS OF ANTIPSYCHOTIC MEDICATION

We calculated haloperidol equivalents for antipsychotic medication dosage for each patient using Expert Consensus Panel guidelines.50 There was no difference in overall antipsychotic burden between the HNS group and the LNS group (\(t = 0.58, P = .56\)). Furthermore, we found no significant correlations between medication dosage and any measures of acquisition, training, or transfer test phase performance. These results suggest that antipsychotic burden is unlikely to account for our findings; however, we cannot rule out an effect of antipsychotic medication on performance that might only be observed by studying non-medicated patients.

COMPUTATIONAL MODELING

The goal of computational modeling was to provide quantitative fits of the overall pattern of acquisition and transfer test phase data by each of 3 models (Appendix, eFigure 1, and eFigure 2). Figure 4B and D show that the best-fitting model reproduces the central features of the data in both training and transfer test phases, including better learning from gain than from loss avoidance (Figure 4B) and preference for frequent winners over frequent loss avoiders at the transfer test phase in the HC group (Figure 4D).

Both of these effects are severely attenuated in the HNS group. The simple actor-critic model was insufficient for the HC group because it captured neither (1) more robust acquisition for winners vs loss avoiders (Figure 4A) nor (2) the observed robust preference for winners over loss avoiders at the transfer test phase. The pure Q-learning model could not account for the observed preference of frequent loss avoiders (FLA in the figures) compared with infrequent winners (IW in the figures) across all groups because infrequent winners have higher expected value (Figure 5B). The critical results are that the hybrid actor-critic–Q-learning model provided the best overall fit to the data and that the HNS group differed from the HC group and the LNS group specifically by demonstrating a reduced Q-learning component.

We tested whether the fitted parameter values from the hybrid model differed by group using ANOVA. We found a main effect of group for the mixing parameter \(c\) (Figure 5A) \((F_{2,67} = 3.8, P = .03)\), indicating a significant difference between groups in the degree to which the Q-value component influenced choices. Follow-up analyses revealed significantly lower contribution of Q values for the HNS group compared with the HC group \((t = 2.77, P = .008)\), as well as a trend in the comparison of the LNS group with the HC group \((t = 1.70, P = .09)\). As shown in Figure 5A, the HC group data were characterized by greater influence of...
Q learning than actor-critic learning, whereas the HNS group showed the opposite pattern.

These results provide insight into the origins of avolition and anhedonia in schizophrenia. First, patients with the most severe negative symptoms demonstrate deficits in learning from rewarding outcomes. This deficit is not a manifestation of a general learning impairment because the HNS group performed at levels similar to those of the HC group when learning to avoid losses. Second, in the transfer test phase, the HNS group did not show a preference for a frequently rewarded stimulus over a frequent loss avoider; that is, they were less able to take expected reward values into account during decision making; therefore, decisions were based on stimulus-response weights learned from prior PEs.

This is an RL formula for avolition: patients are better able to learn actions that lead to the avoidance of punishing outcomes than they are to learn actions that lead to positive outcomes. This pattern of data suggests that negative symptoms are not associated with reduced learning from positive PEs per se, as previously suggested, but rather with impairment in the representation of positive expected value to guide decisions. This conclusion is consistent with other data suggesting that negative symptoms are associated with deficits in reward-based tasks that depend on prefrontal or orbitofrontal cortical function.

Other findings from our group are consistent with the results reported herein. Waltz et al\(^{57}\) reported that patients with schizophrenia showed impaired learning from frequently rewarded stimuli but showed intact avoidance of infrequently rewarded stimuli. In a reanalysis of that data set stimulated by the present findings, it was clear that patients in the HNS group drove the impaired reward learning effect (Appendix, eFigure 1, and eFigure 2). Furthermore, in functional magnetic resonance imaging, Waltz et al\(^{57}\) showed intact modulation of blood
Most work investigating PE signaling in schizophrenia has focused on the possibility that aberrant positive PEs may underlie positive symptoms.\textsuperscript{18,19,58} Our focus is different, and the present design is not optimal for detecting aberrant positive PEs. Therefore, our results do not contradict prior studies but rather suggest that PE-driven RL models may also offer a means of understanding negative symptoms.

The origin of negative symptoms remains a major puzzle. By definition, such symptoms are the absence of normal function. Yet, such an absence must implicate the presence of an underlying causal mechanism. Our data suggest that patients with HNS fail to represent the relative value of different rewards when making decisions, while avoiding losses and punishing outcomes. This is an RL formula for avolition, likely resulting in a narrowing of patients' behavioral repertoires and a failure to activate behavior to accomplish goals.

Submitted for Publication: February 2, 2011; final revision received July 29, 2011; accepted August 4, 2011.

Correspondence: James M. Gold, PhD, Department of Psychiatry, Maryland Psychiatric Research Center, University of Maryland School of Medicine, PO Box 21247, Baltimore, MD 21228 (jgold@mprc.umaryland.edu).

Financial Disclosure: Dr Gold receives royalty payments from the Brief Assessment of Cognition in Schizophrenia and has consulted for Merck, Pfizer, Solvay, GlaxoSmithKline, and AstraZeneca.

Funding/Support: This work was supported by grant R01 MH080066 from the National Institute of Mental Health.

Role of the Sponsors: The funding organization had no role in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; or in the preparation, review, or approval of the manuscript.

Previous Presentation: This study was presented at the 13th International Congress on Schizophrenia Research; April 5, 2011; Colorado Springs, Colorado.

Online-Only Material: The Appendix, eFigure 1, and eFigure 2 are available at http://www.archgenpsychiatry.com.

Additional Contributions: Sharon August, MA; Leeka Hubzin, MA; Jacqueline Kiwanuka, MBA; and Dhivya Pahwa, BA, contributed to the study.

REFERENCES


41. Andreasen NC. *The Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City: University of Iowa; 1984.


