# Selective Reinforcement Learning Deficits in Schizophrenia Support Predictions from Computational Models of Striatal-Cortical Dysfunction

James A. Waltz, Michael J. Frank, Benjamin M. Robinson, and James M. Gold

**Background:** Rewards and punishments may make distinct contributions to learning via separate striatal-cortical pathways. We investigated whether fronto-striatal dysfunction in schizophrenia (SZ) is characterized by selective impairment in either reward- (Go) or punishment-driven (NoGo) learning.

**Methods:** We administered two versions of a probabilistic selection task to 40 schizophrenia patients and 31 control subjects, using difficult to verbalize stimuli (experiment 1) and nameable objects (experiment 2). In an acquisition phase, participants learned to choose between three different stimulus pairs (AB, CD, EF) presented in random order, based on probabilistic feedback (80%, 70%, 60%). We used analyses of variance (ANOVAs) to assess the effects of group and reinforcement probability on two measures of contingency learning. To characterize the preference of subjects for choosing the most rewarded stimulus and avoiding the most punished stimulus, we subsequently tested participants with novel pairs of stimuli involving either A or B, providing no feedback.

**Results:** Control subjects demonstrated superior performance during the first 40 acquisition trials in each of the 80% and 70% conditions versus the 60% condition; patients showed similarly impaired (<60%) performance in all three conditions. In novel test pairs, patients showed decreased preference for the most rewarded stimulus (A; $t = 2.674$; $p = .01$). Patients were unimpaired at avoiding the most negative stimulus (B; $t = .737$).

**Conclusions:** The results of these experiments provide additional evidence for the presence of deficits in reinforcement learning in SZ, suggesting that reward-driven learning may be more profoundly impaired than punishment-driven learning.

**Key Words:** Basal ganglia, dopamine, orbitofrontal, prefrontal, reinforcement, schizophrenia

Cognitive deficits are widely recognized as central features of schizophrenia (SZ) (Barch 2005; Wilk *et al.* 2005). Of the impairments documented in the literature, deficits involving the use of feedback to guide decision making and learning are highly reliable and sometimes clinically dramatic. Patients' poor performance on many of these tasks like the Wisconsin Card Sort Test (WCST) (Goldberg *et al.* 1987) and conditional associative learning paradigms (Gold *et al.* 2000) is often interpreted as evidence of dysfunction in either dorsolateral regions of prefrontal cortex (PFC) (Weinberger *et al.* 1986) or lateral and medial areas of ventral prefrontal cortex (Boettiger and D'Esposito 2005), also called orbitofrontal cortex (OFC).

In contrast, several (but not all) studies of procedural, or habit, learning (Keri *et al.* 2000, 2005; Weickert *et al.* 2002) have documented surprisingly normal learning among SZ patients. These tasks also employ feedback to guide learning but tend to involve gradual learning of difficult-to-discern probabilistic response-outcome relationships. Both functional imaging and studies of patient populations such as Parkinson disease (PD) suggest that the basal ganglia (BG) play a critical role in this gradual learning of stimulus-response mappings (Knowlton *et al.* 1996; Seger and Cincotta 2005).

From the Maryland Psychiatric Research Center (JAW, BMR, JMG), University of Maryland School of Medicine, Baltimore, Maryland; and Department of Psychology (MJF), University of Arizona, Tucson, Arizona.

Address reprint requests to James A. Waltz, Ph.D., University of Maryland School of Medicine, Maryland Psychiatric Research Center, PO Box 21247, Baltimore, MD 21228; E-mail: jwaltz@mprc.umaryland.edu.

Received June 14, 2006; revised September 20, 2006; accepted September 25, 2006.

Explaining the differential impairment of these learning processes in schizophrenia is difficult, given the evidence that brain dopamine (DA) systems are known to play a critical role in both PFC-mediated and BG-dependent reinforcement-learning processes. One possible explanation for the relative sparing of habit learning in SZ is that some DA pathways in the BG are largely intact. To investigate this question, we adopted the experimental methods and computational framework of Frank *et al.* (2004), who examined learning performance in a group of PD patients studied both on and off L-Dopa. Frank *et al.* (2004) used a probabilistic stimulus selection (PSS) task, where subjects are initially presented with three different stimulus pairs (AB, CD, EF) and have to learn to choose the most frequently reinforced stimulus from each pair using probabilistic feedback (Figure 1). After achieving the learning criterion in this acquisition phase, subjects are then presented with the original stimuli in novel pairings in a postacquisition test phase. This design provides a means of studying the contributions of positive and negative feedback to probabilistic learning, in that it enables the assessment of whether subjects have a bias for choosing frequently reinforced stimuli or for avoiding frequently punished stimuli.

Frank *et al.* (2004) demonstrated that unmedicated PD patients showed considerable impairment in learning driven by positive feedback when compared with their performance in the medicated state. Importantly, their learning driven by negative feedback was superior to that in the medicated state. These results were interpreted in the context of computational models of reward-based learning (Frank 2005; Frank *et al.* 2001) that formalize ideas about the role of dopaminergic signaling in the BG. These signals are thought to communicate information about reward contingencies in the environment that guide action selection and learning. A degree of functional segregation characterizes pathways in the BG, such that activity in the direct pathway sends a "Go" signal to facilitate the execution of a

0006-3223/07/$32.00
doi:10.1016/j.biopsych.2006.09.042

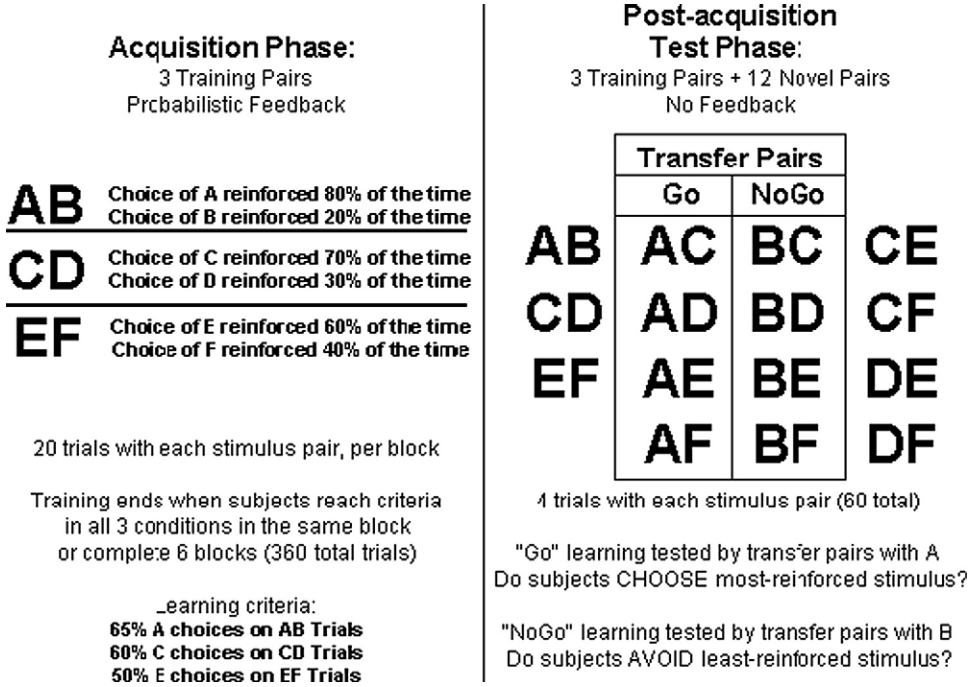


**Figure 1.** The probabilistic stimulus selection (PSS) task. The task consists of two phases. During an acquisition phase, subjects are presented with three training pairs and instructed to identify which stimulus from each pair is more frequently reinforced. In AB trials, for example, a choice of stimulus A leads to positive feedback in 80% of trials, whereas a B choice is reinforced on the remaining 20%. Learning the most frequently rewarded stimulus in each pair can be accomplished either by learning that one of the stimuli leads to positive feedback or that the other leads to negative feedback (or both). Subjects are told to choose that stimulus as often as possible. Once subjects reach criterion on all three training pairs or complete 360 total trials, they proceed to a postacquisition test phase, during which they are presented with four trials each of the three training pairs, along with 12 new pairs created from all unused combinations of the training stimuli. The eight new stimulus pairs involving A and B are called the transfer pairs and are used to gauge Go and NoGo learning. If positive feedback was more effective, they should reliably choose stimulus A in all novel test pairs in which it is present; if they learned more from negative feedback, they should avoid stimulus B. PSS, probabilistic stimulus selection.

response considered in cortex, whereas activity in the indirect pathway sends a "NoGo" signal to suppress inappropriate responses (Figure 2) (Centonze *et al.* 2001; Nishi *et al.* 1997). Furthermore, dopaminergic innervation of these pathways is thought to be relatively distinct, such that the direct pathway is excited via D1 receptors by bursting activity in dopamine neurons, while the indirect pathway is tonically inhibited via D2 receptors. Phasic DA bursts are thought to support Go learning to reinforce rewarding choices by enhancing neural activity and plasticity in the direct (D1) pathway following reinforcement and enhancing inhibition of the indirect (D2) pathway. Transient cessations of DA cell firing, following negative feedback, are thought to have the opposite effect: they release inhibition of the indirect pathway and cause reductions of activity in the direct pathway, thereby supporting NoGo learning to avoid unrewarding choices (Frank 2005; O'Reilly and Frank 2006). These authors concluded that in unmedicated PD patients, DA depletion attenuates the impact of DA bursts. In medicated PD patients, the impact of DA "dips" is attenuated due to overall increased levels of synaptic DA.

While learning at the level of the basal ganglia is thought to occur on a gradual time scale, Go and NoGo signals emanating from the BG are hypothesized to impact the rapid learning of changing reinforcement contingencies in the frontal cortex via parallel striatal-cortical circuits by updating working memory (WM) representations required for representing differences in relative magnitude of reinforcement online (Frank *et al.* 2001; O'Reilly and Frank 2006). This idea extends earlier computational work emphasizing the role of phasic DA in driving the updating of PFC WM representations (Braver and Cohen 2000; Cohen *et al.* 1996). The idea that OFC figures critically in the online representation of reward and punishment magnitudes and thus subserves a kind of working memory is supported by recent evidence (Rolls *et al.* 2003; Schoenbaum and Roesch 2005). Simulations by Frank and Claus (2006) have shown that models capable of instantaneously updating WM representations of reward value in OFC and using them to bias behavior via efferent projections to the BG and motor cortical areas show rapid acquisition of probabilistic contingencies, whereas models with OFC damage exhibit much slower learning because they can only acquire probabilistic contingencies via changes in synaptic weights in the BG.

### Relevance of Dopamine System Function Models to SZ

This framework has the potential to offer a differentiated account of feedback-driven learning deficits in SZ. Whereas PD involves mainly BG hypofunction brought on by dopamine depletion, SZ may be characterized by DA dysfunction in both PFC and the BG. While the severity and consequences of PFC hypofunction in schizophrenia appear to be profound (Weinberger 1987; Weinberger and Berman 1988), BG dysfunction in schizophrenia may be more mild, based on findings of relatively intact procedural learning (Keri *et al.* 2005; Kern *et al.* 1997; Weickert *et al.* 2002).

We tested three specific hypotheses by applying the paradigm used by Frank *et al.* (2004) in their study of PD patients. We