Impaired Implicit Learning in Schizophrenia

William P. Horan
University of California, Los Angeles

Michael F. Green
University of California, Los Angeles, and Veterans Affairs Greater Los Angeles Healthcare System

Barbara J. Knowlton and Jonathan K. Wynn
University of California, Los Angeles

Jim Mintz
University of Texas Health Science Center at San Antonio

Keith H. Nuechterlein
University of California, Los Angeles

Schizophrenia patients consistently show deficits on tasks of explicit learning and memory. In contrast, their performance on implicit processing tasks often appears to be relatively intact, although most studies have focused on implicit learning of motor skills. This study evaluated implicit learning in 59 medicated schizophrenia outpatients and 43 controls using 2 different cognitive skill tasks. Participants completed a probabilistic classification task to assess procedural habit learning and an artificial grammar task to assess incidental learning of complex rule-based knowledge, as well as an explicit verbal learning and memory task. In addition to performing worse than controls on the explicit learning task, patients showed worse overall performance on the probabilistic classification task, which involves gradual learning through trial-by-trial performance feedback. However, patients and controls showed similar levels of learning on the artificial grammar task, suggesting a preserved ability to acquire complex rule-based knowledge in the absence of performance feedback. Discussion focuses on possible explanations for schizophrenia patients’ poor probabilistic classification task performance.

Keywords: schizophrenia, implicit learning, neurocognition, habit learning

Schizophrenia patients consistently demonstrate substantial impairment on neurocognitive tests of explicit learning and memory (Aleman, Bijman, de Haan, & Kahn, 1999; Heinrichs & Zakzanis, 1998), which require effortful conscious awareness during information acquisition and retrieval. These impairments have considerable functional relevance, as they are associated with how well patients function in the community and how much they benefit from psychosocial rehabilitation programs (Green, 1996, Green, Kern, Braff, & Mintz, 2000). In contrast, the smaller literature on implicit learning and memory in schizophrenia shows much more variability. Implicit learning refers to skill, habit, or knowledge acquisition that occurs automatically and outside of conscious awareness, whereas implicit memory refers to information retrieval that occurs outside of conscious awareness (Squire & Zola, 1996). The majority of implicit processing studies in schizophrenia have examined motor learning tasks, such as the pursuit rotor task or mirror drawing task (e.g., Kern, Green, & Wallace, 1997; Sponheim, Steele, & McGuire, 2004; Wexler et al., 1997). These studies tend to find intact implicit motor learning in schizophrenia, although impairments in implicit learning have been found in some studies, particularly on the serial reaction time task (see, e.g., Exner, Bousche, Degner, & Irle, 2006; Exner, Weniger, Schmidt-Samoa, & Irle, 2006; Green, Kern, Williams, McGurk, & Kee, 1997; Schwartz, Howard, Howard, Hovaguimian, & Deutsch, 2003; see also Goldberg, Saint-Cyr, & Weinberger, 1990; Gras-Vincendon et al., 1994; Michel, Danion, Grange, & Sandner, 1998).

Implicit learning is a multifaceted construct and it can be roughly divided into motor and cognitive learning, which appear to rely on different neural circuits (Ashby & Maddox, 2005; Keri, 2003; Yin & Knowlton, 2006). To date, implicit learning of cognitive skills has received comparatively less attention in schizophrenia research. The most extensively studied cognitive implicit learning tasks in the basic cognitive science literature involve perceptual category learning, such as the “weather prediction” or probabilistic classification task (PC task; Knowlton, Mangels, & Squire, 1996) and artificial grammar tasks (AG tasks; Reber, 1967). In the PC task, participants gradually learn to predict the weather (will it “rain” or “shine”?) across a number of trials based on associations between stimuli (geometric forms) that are independently and probabilistically related to the weather...
outcome (rather than a deterministic, one-to-one mapping between each stimulus combination and outcome). After each trial, the participant is informed that his choice is right or wrong, but no other feedback is provided. Healthy participants typically show significant classification improvements across learning trials but, because of the probabilistic nature of this task, are unable to explicitly state the rules that govern their predictions.

The AG task also involves category learning across a series of trials, yet it differs from the PC task in several respects. In the learning stage of the task, participants are shown a series of exemplars of grammatical letter strings that are based on a complex, finite-state grammar system. In a separate testing stage of the task, participants are asked to classify a new set of letter strings as either grammatical or not grammatical, with no feedback provided about the accuracy of their choices. Healthy participants typically perform significantly above-chance levels in their grammaticality classifications but are unable to explicitly state the rules that govern their responses because of the complexity of the underlying grammar system.

The PC and AG category learning tasks appear to rely on different neural circuitry. For example, successful performance on PC tasks typically does not depend on the medial temporal lobe explicit memory system, although it can under certain circumstances (e.g., (Foerde, Knowlton, & Poldrack, 2006). Amnesic and Alzheimer’s patients with medial temporal lobe damage have typically been found to perform normally on both the PC task (e.g., Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, & Gluck, 1994; but see Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004) and the AG task (e.g., Knowlton & Squire, 1996; Reber, Martinez, & Weintraub, 2003). In contrast, patients with Parkinson’s or Huntington’s disease, who exhibit abnormal striatal dopaminergic functioning, show impaired performance on the PC task yet preserved AG task performance (Eldridge, Masterman, & Knowlton, 2002; Knowlton, Mangels, & Squire, 1996; Knowlton et al., 1994; Knowlton et al., 1996; Smith, Siegert, McDowall, & Abernethy, 2001). In conjunction with human neuroimaging studies (e.g., Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Moody, Bookheimer, Vanek, & Knowlton, 2004; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Yin & Knowlton, 2006), these findings suggest that an intact neostriatal habit learning system is critical for PC task performance but not for the AG task.

Only a handful of studies have examined these category learning tasks in schizophrenia. For the PC task, two studies reported that patients demonstrated worse overall percent correct scores than healthy controls, although the slopes of the learning curves during the task did not significantly differ between groups (Keri et al., 2005; Weickert et al., 2002). Another study reported performance deficits only in patients taking typical antipsychotic medications, which have relatively high dopamine D2 receptor affinities, whereas patients taking atypical antipsychotics did not significantly differ from controls (Beninger et al., 2003). Finally, one study reported intact performance in schizophrenia (Keri et al., 2000) but used an idiosyncratic administration procedure that makes comparison with the other studies difficult.

For the AG task, we are aware of only two studies that examined performance in schizophrenia patients. Danion and colleagues (Danion, Meulemans, Kauffmann-Muller, & Vermaat, 2001) found that schizophrenia patients and healthy controls both significantly exceeded chance levels in their accuracy judgments and did not significantly differ from each other. A subsequent study also reported intact AG performance in a small sample of 10 schizophrenia patients (Hsieh et al., 2004).

There has been some debate among cognitive scientists about whether the learning that occurs on the AG task truly reflects deep, abstract rule-based learning or instead reflects superficial, exemplar-specific learning. On average, grammatical test strings are likely to bear greater superficial similarity (e.g., in terms of specific bigram or trigram chunks that also occur during training trials) to the learning strings than do the nongrammatical test strings. This superficial similarity between learning and testing strings creates an interpretative confound in traditional AG tasks by providing an alternative learning strategy that does not rely on deep, rule-based knowledge.

To address this confound, researchers have used a “balanced chunk strength” design, in which grammatical and nongrammatical test strings are intentionally matched for both high and low chunk strengths (Knowlton & Squire, 1994, 1996). Chunk strength refers to the number of times bigram or trigram chunks within a test string are repeated across the training strings. This design enables investigators to directly examine whether participants’ use complex, rule-based knowledge, beyond information about the superficial characteristics of test strings in making their grammaticality judgments. Research using this experimental design in healthy participants indicates that, although grammaticality judgments are to some extent influenced by chunk strength, they are indeed influenced by abstract rule-based knowledge as well (Chang & Knowlton, 2004; Knowlton & Squire, 1996). In the schizophrenia research literature, Danion and colleagues (Danion et al., 2001) used grammatical and nongrammatical test strings that were matched for mean levels of chunk strength. In the present study, chunk strength was manipulated as an independent variable using a balanced chunk strength design to allow us to directly evaluate whether patients use abstract, rule-based knowledge in their grammaticality judgments.

The present study examined the performance of stabilized schizophrenia outpatients and healthy controls on the PC and AG tasks, as well as measures of general verbal ability and explicit learning. By examining the performance of patients on two different implicit learning tasks, we will be able to better identify those aspects of implicit learning that may be impaired in schizophrenia. On the basis of existing studies, we predicted that schizophrenia patients would demonstrate worse performance than controls on the explicit learning and memory tasks and the PC task but that the groups would show comparable levels of learning on the AG task.

Method

Participants

Participants included 59 schizophrenia outpatients and 43 nonpatient controls. Patients met criteria for schizophrenia from the Structured Clinical Interview for DSM–IV (SCID; First, Gibbon, Spitzer, & Williams, 1996). Diagnostic interviewers were trained to a minimum kappa of 0.75 for rating psychotic and mood symptoms by the Treatment Unit of the Veterans Integrated Service Network 22, Mental Illness Research, Education, and Clinical Center (Los Angeles, CA). Patients were excluded for current substance use disorders (within the past 6 months) but not excluded for only a history of substance use disorder. All patients
were receiving various antipsychotic medications at clinically determined dosages (n = 50 for atypical only; n = 3 for typical only; n = 6 for both). Four patients were prescribed benzodiazepines, and these patients were asked to refrain from taking these medications on the day of testing (i.e., at least 12 hr before assessment). Nine patients were prescribed anticholinergic medications.

Nonpatient controls were recruited through newspaper advertisements and flyers posted in the local community. Controls were screened with the SCID for Axis I and Axis II disorders (First, Spitzer, Gibbon, Williams, & Benjamin, 1994) and were excluded if they met criteria for any psychotic disorder; bipolar mood disorder; recurrent depression; substance dependence; or paranoid, schizotypal, schizoid, avoidant, or borderline personality disorder. Controls were also excluded for current substance abuse (within the past 6 months) or if there was any evidence (according to participant report) of a history of psychotic disorder among their first-degree relatives. Additional exclusion criteria for all patients and controls included an age less than 18 or over 55 years, identifiable neurological disorder, mental retardation, or seizure disorder.

Clinical Ratings

Brief Psychiatric Rating Scale (BPRS)

For all patients, a trained rater used the expanded 24-item University of California, Los Angeles, version of the BPRS (Lukoff, Nuechterlein, & Ventura, 1986; Overall & Gorham, 1962) to rate all psychiatric symptoms during the previous month. Each item is rated on a scale ranging from 1 (Not reported) from 7 (Very severe). BPRS raters achieved a median intraclass correlation coefficient of .80 or higher across all items compared with the criterion ratings (Ventura, Green, Shaner, & Liberman, 1993). From this version of the BPRS, five empirically derived subscale scores (based on the mean of items constituting the scale; Guy, 1976) and a 24-item total score are derived.

Scale for the Assessment of Negative Symptoms (SANS)

Negative symptoms during the preceding month were evaluated using the SANS (Andreasen, 1984). Four SANS global scales were used in the present study: Affective Flattening, Alogia, Anhedonia–Asociality, and Avolition–Apathy. The SANS Attention scale was not included in the present analyses, given findings suggesting that this scale is not conceptually related to the negative symptom complex (e.g., Blanchard & Cohen, 2006).

Measures

Wechsler Test of Adult Reading (WTAR; Wechsler, 2001)

This is a task of general word knowledge that is frequently used as an estimate of premorbid verbal ability. Participants are handed a written list of 50 words and are asked to read aloud all of the words to the best of their ability. A total score is calculated from the number of correctly pronounced words.

Hopkins Verbal Learning Test—Revised (HVLT)

In the HVLT (Benedict, Schretlen, Groninger, & Brandt, 1998; Brandt & Benedict, 2001), a list of 12 words drawn from three semantic categories is presented across three learning trials with recall assessed after each trial. After the third trial, a new list of 12 words is presented and recall assessed. On the HVLT, the short delay trial is administered immediately after the second (distractor) word list is administered. Then the long delay trial is administered 20 minutes later. After the long-delay recall, a 24-item recognition test is administered. Dependent measures selected for the present study were the total number of correct responses for each of the three learning trials and an overall learning slope.

Probabilistic Classification Task

Procedure. This computerized task was identical to that procedure described by Knowlton et al. (1996). Participants were instructed that they would see various combinations of four different cue cards (depicting circles, diamonds, squares, or triangles) on each trial and that their task was to decide whether each particular combination of cues they were shown predicted either rain or sunshine. Each cue card was associated with one of two outcomes (rain vs. sunshine) with a fixed probability. One, two, or three cue cards appeared on each trial, and the two outcomes occurred equally often. Thus, there were 14 possible cue combinations. As shown in Table 1, the overall probability of each outcome (rain or sunshine) on a given trial was calculated according to the conditional probabilities for each of the presented cue cards.

Each participant received one of two different randomized sequences of cue combinations (Sequence A or Sequence B), that were each constructed with the constraint that the same cue card combination could not appear twice in succession, and each outcome occurred no more than five times in a row. Participants were told that at first they would be guessing but that they would gradually become better at deciding which cues predicted sunshine or rain.

At the start of each trial, the cues appeared on the computer screen for 5 s. The participant indicated his or her choice by pressing either the key labeled “sunshine” or the key labeled “rain” on the computer keyboard. If the response was correct, a high-pitched tone sounded, a smiling face appeared at the right of the screen, and a vertical scoring bar at the right of the screen, initially set at 600, incremented 1 point. If the response was incorrect, a low-pitched tone sounded, a frowning face appeared at the right side of the screen, and the vertical bar at the right of the screen decremented by 1 point. Then, the icon (sun or rain) corresponding to the correct answer appeared on the screen above the cues for 2 s. If no response was made within 3 s, a prompt appeared on the screen asking for a response. After 5 s without a response, the trial was terminated, the low tone sounded, and the correct answer appeared above the cues for 2 s (these missed trials were not scored). Participants were allowed a short break (<1 min) after 50 trials. After a total of 100 trials were completed, they were administered an eight-item, four-alternative, multiple-choice explicit memory task that asked about the nature of the cues and feedback, the layout of the screen, and the testing procedure (e.g., How many squares appeared on the card with squares?).

Scoring. A response was considered to be correct on a particular trial if the outcome selected was the outcome more strongly associated with the cue combination that appeared on that trial. Because of the probabilistic nature of the task, a cue combination
was sometimes followed by the less strongly associated outcome. Thus, a response could be scored as correct on a particular trial, even though the feedback indicated that the response was incorrect. In this way, the percent correct score reflected how well participants learned the cue–outcome associations. Because the two outcomes occurred equally often, chance performance was 50% correct. The data were not analyzed for trials on which the two outcomes were equally associated with the cue combination and for which there was, therefore, no correct answer. These trials composed 10.7% of all the trials. Percent correct was calculated for each block of 20 trials. In addition, an overall learning slope and a total score on the explicit memory task were calculated for each participant.

AG Task

This procedure followed the exact specifications of Chang and Knowlton (2004), consisting of a training phase and a testing phase. Before training, participants were told that they were going to be shown a series of cards containing letter strings that they were to memorize. There was no mention of the AG rule system.

Training. Participants were presented with a series of 23 grammatical strings that was presented twice. All strings were generated from a finite-state Markovian rule system consisting only of the letters X, J, V, and T and were thus grammatical (see Figure 1). Grammatical letter strings were generated by starting at the “IN” arrow and picking up a letter while traversing from one state to the next until an “OUT” arrow is reached. All letter strings varied in length from two letters to six letters and were printed in 100-point black Times font on white 5/8-in. (12.7 × 20.32-cm) index cards. Each index card containing one string was presented for 3 s. After each presentation, participants were asked to reproduce the string on paper. If a participant correctly reproduced the string, the next card was shown. If the participant did not correctly reproduce the string, he or she was shown the same card again for 3 s and was asked to reproduce it again. If the participant still could not reproduce the string, this process was repeated. We allowed participants up to three trials to reproduce each string to ensure that all letter strings were attended during encoding. After three presentations and three attempts to reproduce the string, the experimenter moved on to present the next string.

There was a 5-min delay between the training and testing phases. To discourage active rehearsal of the training items, we administered the Raven Progressive Matrices Task (Raven, 1941) during the delay. This is a nonverbal reasoning task that would be unlikely to interfere with knowledge of the AG.

Testing. After the 5-min delay, participants were given the following instructions.

Table 1

<table>
<thead>
<tr>
<th>Cue pattern</th>
<th>P (cue combination appearing)</th>
<th>P (feedback indicating outcome 1 is correct)</th>
<th>Outcome predicted by structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 0 0 1</td>
<td>.133</td>
<td>.15</td>
<td>2</td>
</tr>
<tr>
<td>2 0 0 1 0</td>
<td>.087</td>
<td>.385</td>
<td>2</td>
</tr>
<tr>
<td>3 0 0 1 1</td>
<td>.080</td>
<td>.083</td>
<td>2</td>
</tr>
<tr>
<td>4 0 1 0 0</td>
<td>.087</td>
<td>.615</td>
<td>1</td>
</tr>
<tr>
<td>5 0 1 0 1</td>
<td>.067</td>
<td>.200</td>
<td>1</td>
</tr>
<tr>
<td>6 0 1 1 0</td>
<td>.04</td>
<td>.500</td>
<td>Neither</td>
</tr>
<tr>
<td>7 0 1 1 1</td>
<td>.047</td>
<td>.143</td>
<td>2</td>
</tr>
<tr>
<td>8 1 0 0 0</td>
<td>.133</td>
<td>.850</td>
<td>1</td>
</tr>
<tr>
<td>9 1 0 0 1</td>
<td>.067</td>
<td>.500</td>
<td>1</td>
</tr>
<tr>
<td>10 1 0 1 0</td>
<td>.067</td>
<td>.800</td>
<td>1</td>
</tr>
<tr>
<td>11 1 0 1 1</td>
<td>.033</td>
<td>.400</td>
<td>2</td>
</tr>
<tr>
<td>12 1 1 0 0</td>
<td>.08</td>
<td>.917</td>
<td>1</td>
</tr>
<tr>
<td>13 1 1 0 1</td>
<td>.033</td>
<td>.600</td>
<td>1</td>
</tr>
<tr>
<td>14 1 1 1 0</td>
<td>.047</td>
<td>.857</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. On any trial, 1 of 14 possible combinations of four cues appeared with the probability indicated [P(cue combination appearing)]. Each cue was associated with one of the two weather outcomes (1 = sunshine vs. 2 = rain) with a fixed probability. Across all 14 combinations of cues, Cue 1 was associated 75% of the time with Outcome 1 and 25% of the time with Outcome 2; Cue 2 was associated 56% of the time with Outcome 1 and 44% of the time with Outcome 2; Cue 3 was associated 44% of the time with Outcome 1 and 56% of the time with Outcome 2; Cue 4 was associated 25% of the time with Outcome 1 and 75% of the time with Outcome 2. Each combination of cues predicted Outcome 1 (Sunshine) with the probability shown [P(feedback indicating that Outcome 1 is correct)] and predicted Outcome 2 (Rain) with a probability of 1 – P.

Figure 1. Finite-state Markovian rule system used to generate grammatical letter strings (Abrams & Reber, 1989). S1–S5 = the five possible states that occur during generation of grammatical letter strings.
The orders of the letters in the previous set of cards were actually determined by a complex set of grammatical rules. You will now be presented with another set of cards for an unlimited amount of time. Your task is to decide whether each item was or was not formed according to the same rules. If the current letter string follows the same rules as the strings you were shown earlier, they should be considered “grammatical.” Otherwise, they should be considered “not grammatical.” Since these grammatical rules are very complex, you should base your judgments on a gut feeling as to whether each test item obeys these rules without trying to figure out the rules. Please determine if the following letter strings are “grammatical” or “not grammatical.”

Participants were shown a series of 32 cards, each of which had a letter string that had not been previously seen. Half of the strings were grammatical, and half were not grammatical (nongrammatical test strings were formed by introducing an error in one position to an otherwise grammatical string). The order of the strings was mixed so that no more than three items of the same chunk strength (either high or low) or grammatical status (grammatical or not grammatical) were shown in succession. Chunk strength was calculated as the average number of times each of the bigrams and trigrams in the string had been presented in the training set (see Knowlton & Squire, 1996). As shown in Table 2, the average chunk strength of high-chunk-strength items was 8.5; the average strength of low-chunk-strength items was 5.6. Participants’ verbal responses to the grammatical status of the letter strings were recorded by the experimenter.

Half of the test items had high chunk strengths, whereas the other half had low chunk strengths. Of the grammatical and nongrammatical test strings, there were an equal number of high- and low-chunk-strength items. From this task, an overall percent correct score was calculated, as well as the number of high- and low-chunk-strength items that were endorsed as grammatical during the testing trials.

**Data Analysis**

Group differences in demographics were evaluated with paired t tests for continuous variables and chi-square tests for categorical variables. Data were missing for 3 patients and 1 control on the PC task. Mean and SD for the nonindependence of observations across trials, and group as a fixed factor. In addition, linear function terms were included as covariates to evaluate the presence of Group × Learning Slope interactions (preliminary analyses indicated that linear effects were much stronger than quadratic effects for the learning slopes). Effect sizes for these analyses are presented as F², which corresponds to the following effect size conventions: small (.02), medium (.15), and large (.35; Cohen, 1988).

On the AG task, group differences in overall percent correct scores were evaluated with independent samples t tests. These were followed by a 2 (group) × 2 (chunk strength) × 2 (grammar) analysis of variance (ANOVA) to evaluate the sensitivity of judgments of grammar and chunk strength. Effect sizes are presented as Cohen’s partial eta squared (ηp²), which corresponds to the following conventions: small (.01), medium (.06), and large (.14; Cohen, 1988). Finally, within each group, associations among measures were evaluated with Pearson correlation coefficients. Associations between performance measures and symptoms were also evaluated within the patient group.

**Results**

**Participant Characteristics and Verbal Ability**

As shown in Table 3, the groups did not significantly differ on age, gender, education, parental education, ethnicity, or marital status. This outpatient schizophrenia sample had mild to moderate levels of symptoms on the BPRS subscales and total scores that are typical for stabilized, community-dwelling samples. Mean ratings on the BPRS subscales (possible ranges = 1–7) were as follows: Anergia, M = 1.8, SD = 0.7; Anxiety/Depression, M = 2.5, SD = 0.9; Thought Disturbance, M = 2.8, SD = 1.1; Activation, M = 1.2, SD = 0.3; and Hostile Suspiciousness, M = 2.1, SD = 0.7. The mean BPRS Total score (based on the sum of the 24 BPRS items, possible range = 24–168) was 46.3 (SD = 9.9). Mean scores on the SANS subscales were as follows: Affective Flattening, M = 1.9, SD = 1.3; Alogia, M = 0.8, SD = 1.0; Apathy/Avolition, M = 2.9, SD = 1.2; and Anhedonia/Asociality, M = 2.5, SD = 1.2. Patients had a mean duration of illness of 17.4 years (SD = 9.3). On the WTAR, scores were significantly lower for patients (M = 30.6,
a linear effect for the learning slope across trials, $F(1, 202) = 393.74, p < .001, F^2 = 1.95$. However, the Group × Slope interaction was not significant, $F(1, 202) = 1.69, p > .05, F^2 = 0.01$, suggesting comparable learning curves across the groups.

**Probabilistic Classification Task**

The number of no-response trials was very low in both groups, with similar levels found in patients ($M = 0.30, SD = 0.79$) and controls ($M = 0.30, SD = 0.62$), $t(96) = 0.11, p > .05, d = 0.02$. Group × Trial Block (in blocks of 20 trials) data are displayed graphically in Figure 3. A mixed-effects regression analysis revealed significant effects for group, $F(1, 96) = 7.12, p < .01, F^2 = 0.07$, indicating lower overall scores for patients. The learning slope across trials demonstrated a significant linear function, $F(1, 390) = 47.08, p < .001, F^2 = 0.12$. However, the Group × Learning Slope interaction was not significant, $F(1, 390) = 1.94, p > .05, F^2 = 0.001$, indicating generally comparable overall learning rates in both groups. In addition, the group effect remained significant with WTAR scores entered as a covariate, $F(1, 95) = 5.95, p < .01, F^2 = 0.06$. The magnitude of the group difference on total percent correct scores between patients ($M = 0.59; SD = 0.11$) and controls ($M = 0.65; SD = 0.14$) was medium ($d = 0.47$). Although the Group × Learning Slope interaction was not statistically significant, the effect size estimates presented in Figure 3 indicate that the magnitude of the between-groups differences tended to be larger during the later trial blocks of the task. A supplemental analysis of group differences within each trial block indicated that patients and controls did not significantly differ during the first two blocks, but did significantly differ during the final three blocks ($ps < .05$).

During the first block of 20 trials, both groups performed better than typically expected on this task. A more detailed analysis of the data indicated that there were some performance differences during early trials depending on the trial sequence received by the

![Figure 2](https://example.com/Figure2.png)

**Figure 2.** Hopkins Verbal Learning Test performance in the schizophrenia and control groups. Error bars represent standard errors. $d =$ standardized effect size for between-groups difference.
participant. For Sequence A \((n = 46)\), one of the cues was strongly linked to the outcome “sun” across the first 10 trials, which resulted in above-chance performance for both groups in the first block of 20 trials. This earlier learning in the first 10 trials was nonsignificantly greater in control participants \((M = 0.65, SD = 0.18)\) than in the patients \((M = 0.56, SD = 0.21)\), \(t(44) = 1.55, p > .05, d = .47\). In contrast, Sequence B \((n = 52)\) did not appear to afford such early learning, with more typical chance-level performance shown during the first 10 trials by both controls \((M = 0.51, SD = 0.16)\) and patients \((M = 0.50, SD = 0.18)\). Despite these early differences in performance on the two trial sequences, the overall pattern of results was similar for participants receiving the different sequences, with no significant Group \(\times\) Slope interactions for Sequence A or B \((ps > .05)\).

On the accompanying explicit memory task, scores were significantly lower for patients \((M = 0.61, SD = 0.11)\) than for controls \((M = 0.67, SD = 0.13)\), \(t(96) = 2.22, p < .05, d = .45\). However, scores for each group were significantly above chance level \((ps < .001)\), suggesting adequate attentiveness to the task in both groups.

Because of the relatively high striatal D\(_2\) receptor blockade associated with typical antipsychotic medications, supplemental analyses examined whether the aforementioned findings were attributable to patients taking these medications. There was no significant difference in total percent correct scores for patients taking typical \((n = 6; M = 0.59, SD = 0.11)\) versus atypical \((n = 50; M = 0.59; SD = 0.13)\) antipsychotics, \(t(54) = 0.01, p > .05, d = 0.002\). We also reran the mixed-effects regression analysis, entering just those patients who were taking only atypical antipsychotics. Fifty of the 56 patients who completed the PC task were taking only atypical antipsychotic medications. Results were highly similar, with significant effects for group, \(F(1, 88) = 6.29, p < .05, F^2 = 0.07\); and learning slope (linear function) across trials, \(F(1, 358) = 51.89, p < .001, F^2 = 0.14\); but a nonsignificant Group \(\times\) Linear Trend interaction, \(F(1, 358) = 0.98, p > .05, F^2 = 0.003\), suggesting a similar pattern among patients taking only atypical antipsychotics. Finally, we reran the mixed-effects regression analysis entering only those patients who were not prescribed anticholinergic medications at the time of testing \((n = 47)\). Again, the results were highly similar to the full analyses, with significant effects for group, \(F(1.87) = 6.05, p < .05, F^2 = 0.07\); and linear trend, \(F(1, 354) = 45.29, p < .001, F^2 = 0.13\); but a nonsignificant Group \(\times\) Linear Trend interaction, \(F(1, 354) = 1.56, p > .05, F^2 = 0.004\), suggesting that the group differences were not solely accounted for by patients taking anticholinergics.

\textbf{AG Task}

During the learning phase, the number of learning trials required was significantly greater for patients \((M = 54.8, SD = 9.3)\) than for controls \((M = 51.3, SD = 4.4)\), \(t(98) = 2.49, p < .05, d = 0.50\). During the testing phase, the total percentages of correct responses were significantly above chance in both patients \((M = 0.56, SD = 0.10)\) and controls \((M = 0.58, SD = 0.10)\), providing evidence consistent with rule-based learning in both groups. These percentages did not significantly differ between
groups, \( t(98) = 1.08, p > .05, d = 0.22 \). In addition, there were no significant differences between patients taking typical (\( M = 0.55, SD = 0.09 \)) versus atypical (\( M = 0.56, SD = 0.10 \)) antipsychotic medications, \( t(55) = 0.30, p > .05, d = 0.08 \).

During the training phase, the schizophrenia group required more learning trials than the controls, raising concerns that the patients’ apparently normal performance might be attributable to higher levels of exposure to the training stimuli. We reran the analyses using a subset of patients (\( n = 37 \)) who were matched to controls on total learning trials by discarding from the analyses those patients who needed more presentations than the mean of the comparison participants plus one standard deviation. Results confirmed the previous analyses, as total percent correct scores in this subset of patients (\( M = 0.97, SD = 0.08 \)) did not significantly differ from controls (\( p > .05 \)). Thus, additional exposure to training strings in a subset of patients does not appear to be the cause of the patient groups’ normal performance on the AG task.

Next, using the entire sample, we evaluated the endorsement frequencies for specific types of test items using a 2 (group) \( \times \) 2 (grammaticality) \( \times \) 2 (chunk strength) ANOVA. As displayed graphically in Figure 4, patients and controls were responsive to both grammaticality and chunk strength in their responses. There were significant main effects of grammatical status, \( F(1, 98) = 64.06, p < .001, \eta_p^2 = .64 \), and chunk strength, \( F(1, 98) = 200.36, p < .001, \eta_p^2 = .67 \), demonstrating that participants more frequently endorsed items that followed the grammatical rules and items with high-chunk strengths. However, the group main effect was nonsignificant, \( F(1, 98) = 1.87, p > .05, \eta_p^2 = .02 \); and there were no significant interactions between group and the other two variables (all \( ps > .05 \)).

Across groups, there was a significant Grammaticality \( \times \) Chunk Strength interaction, \( F(1, 98) = 16.46, p < .001, \eta_p^2 = .14 \). Post hoc tests indicated that chunk strength exerted most of its effect on nongrammatical items. Endorsement frequencies did not significantly differ for high-chunk-strength grammatical versus high-chunk-strength nongrammatical items, \( \eta_p^2 = .39 \), and chunk strength, \( F(1, 98) = 200.36, p < .001, \eta_p^2 = .67 \), demonstrating that participants more frequently endorsed items that followed the grammatical rules and items with high-chunk strengths. However, the group main effect was nonsignificant, \( F(1, 98) = 1.87, p > .05, \eta_p^2 = .02 \); and there were no significant interactions between group and the other two variables (all \( ps > .05 \)).

Although the patients demonstrated significantly worse overall performance than controls in the PC task, one might question whether the patients would achieve normal levels if they were given additional learning trials. The present results do not clearly support this possibility, as the magnitudes of the between-groups differences were generally larger during the later trial blocks. In addition, a recent study of automaticity in schizophrenia by Foerde et al. (Foerde et al., in press) that included 600 PC trials does not support this possibility. In that study, schizophrenia patients and healthy controls completed twelve 50-trial blocks of the PC task (under single-task or dual-task conditions) over three separate training sessions. Controls reached asymptote by the end of the first session and showed little improvement in the subsequent sessions. In contrast, patients showed only slight improvement across the three sessions and never reached the level of performance that controls reached after many more trials than were used in the present study.

### Correlational Analyses

As shown in Table 4, within the schizophrenia group, there were no significant correlations between verbal ability, explicit task variables, and implicit task variables. Higher learning slopes on the PC task significantly correlated with higher hostility ratings, which was not predicted and possibly a chance finding because of the number of correlations examined. Associations with symptoms were otherwise generally nonsignificant and small. Within the control group, there were no significant correlations among the cognitive task variables, with no correlations exceeding .25.

### Discussion

Consistent with a large body of research, schizophrenia patients demonstrated explicit learning and memory impairments on tasks that involved effortful, conscious awareness at the time of knowledge acquisition and retrieval. This was evident in patients’ lower scores on the HVLT, as well as in their lower scores on the explicit memory test in the PC task and their need for more learning trials on the AG task. The lack of association between the explicit and implicit tasks in both groups is consistent with the theoretical distinction between explicit and implicit learning processes (Squire & Zola, 1996).

In contrast to the patients’ pervasive explicit learning deficits, they demonstrated a varied pattern of performance across the cognitive implicit learning tasks. The lack of association between the PC and AG tasks in both groups is consistent with theoretical distinctions between different types of cognitive implicit learning (Ashby & Maddox, 2005; Keri, 2003). Consistent with prior studies of schizophrenia patients taking primarily atypical antipsychotics (Keri et al., 2005; Weickert et al., 2002), patients demonstrated an overall impairment on the PC task, as compared with controls. However, they performed as well as controls on an AG task that used a balanced chunk strength design, which clearly indicated that patients used deep, rule-based knowledge (in addition to superficial exemplar-based learning) to perform this task. This finding converges with prior AG studies suggesting that at least some aspects of cognitive implicit learning may be preserved in schizophrenia (Danion et al., 2001; Hsieh et al., 2004), similar to the pattern of cognitive implicit learning found in Parkinson’s and Huntington’s patients (Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, et al., 1996).

### Results

Results were similar when the groups were compared on \( d' \), a related signal detection discrimination index that accounts for incorrect detection rates, with patients (\( M = 0.43, SD = 0.49 \)) not significantly differing from controls (\( M = 0.54, SD = 0.52 \)), \( t(98) = 1.12, p > .05 \).

2 The full correlation matrix is available from the authors on request.
Why might the schizophrenia patients have demonstrated worse performance than controls on the PC task but not on the AG task? Although these tasks differ in several respects (see Smith et al., 2001), the most salient difference may be that the PC involves gradually learning to associate stimuli with behavioral responses through probabilistically determined, trial-by-trial feedback, whereas no feedback is provided in the AG task. An interesting set of studies strongly suggests that the impaired PC performance shown by Parkinson’s patients (in the context of normal AG performance) reflects impaired ability to use basal ganglia-mediated feedback to guide learning of stimulus–response associations, rather than differences in the format or complexity of these tasks. For example, Shohamy and colleagues (Shohamy, Myers, Onlaor, & Gluck, 2004) modified the PC task to eliminate the feedback aspect of the original procedure and found that Parkinson’s patients performed normally on the modified no-feedback version of the task but worse than controls on the original version. Similarly, Smith et al. (2001) modified the original AG task to incorporate trial-by-trial feedback and found that Parkinson’s patients demonstrated impaired performance on the modified feedback version of the task. Smith and McDowall (2001) proposed that patients with pathology of the striatum differ from controls in their ability to incrementally learn complex categorization rules based on trial-by-trial feedback. This explanation is consistent with evidence that dopaminergic transmission in the basal ganglia plays a crucial role in reinforcement learning and reward-based decision-making by generating reward prediction errors (Frank & Claus, 2006; Packard & Knowlton, 2002; Schultz, Dayan, & Montague, 1997). In line with these findings, Parkinson’s patients show abnormal error-related negativity (Falkensetin, 2001), an electrophysiological index of error monitoring.

One speculative, post hoc explanation for the present findings is that the poor PC performance of schizophrenia patients reflects impaired feedback-based learning, perhaps associated with striatal pathology. Striatal pathology may be a core pathophysiological feature of this disorder (Chatterjee et al., 1995; Corson, Nopoulos, Andreasen, Heckel, & Arndt, 1999; Gur et al., 1998). We previously found abnormal reward processing in schizophrenia using the Iowa gambling task, a category learning task that depends on different types of reward or punishment feedback (Shurman, Horan, & Nuechterlein, 2005). In addition, schizophrenia patients show abnormal error-related negativity (Morris, Yee, & Nuechterlein, 2006) and, in fMRI studies, decreased striatal activation during reward processing (Juckel et al., 2006) and motor implicit learning (Reiss et al., in press). The hypothesis of impaired feedback-based implicit learning in schizophrenia could be directly evaluated in future research by administering feedback and nonfeedback versions of the PC and AG tasks (e.g., Shohamy et al., 2004; Smith et al., 2001). Further research on basal ganglia-mediated reward processing could also help illuminate the mechanisms of schizo-

---

Table 4

<table>
<thead>
<tr>
<th>Scale</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. WTAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. HVLT: Slope</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. PC: Task slope</td>
<td>.04</td>
<td>.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. AG task: Percent correct</td>
<td>.17</td>
<td>-.04</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>5. BPRS: Thought disturbance</td>
<td>.05</td>
<td>.30</td>
<td>.08</td>
<td>.25</td>
</tr>
<tr>
<td>6. BPRS: Anxiety/Depression</td>
<td>.06</td>
<td>.21</td>
<td>-.01</td>
<td>.28</td>
</tr>
<tr>
<td>7. BPRS: Anergia</td>
<td>.30</td>
<td>-.16</td>
<td>.08</td>
<td>.20</td>
</tr>
<tr>
<td>8. BPRS: Hostility</td>
<td>-.11</td>
<td>.12</td>
<td>.33*</td>
<td>-.09</td>
</tr>
<tr>
<td>9. BPRS: Activation</td>
<td>.10</td>
<td>.32</td>
<td>.08</td>
<td>.13</td>
</tr>
<tr>
<td>10. BPRS: Total</td>
<td>.23</td>
<td>.15</td>
<td>.08</td>
<td>.30</td>
</tr>
<tr>
<td>11. SANS: Affective flattening</td>
<td>.12</td>
<td>.02</td>
<td>.05</td>
<td>.02</td>
</tr>
<tr>
<td>12. SANS: Alogia</td>
<td>.19</td>
<td>.12</td>
<td>.15</td>
<td>.01</td>
</tr>
<tr>
<td>13. SANS: Avolition/Apathy</td>
<td>-.14</td>
<td>-.02</td>
<td>-.17</td>
<td>.04</td>
</tr>
<tr>
<td>14. SANS: Anhedonia/Asocial</td>
<td>-.23</td>
<td>-.03</td>
<td>.06</td>
<td>.10</td>
</tr>
</tbody>
</table>

Note. WTAR: Wechsler Test of Adult Reading; HVLT: Hopkins Verbal Learning Test; PC: probabilistic classification; AG: artificial grammar; BPRS: Brief Psychiatric Rating Scale; SANS: Scale for the Assessment of Negative Symptoms.

*p < .05.
phrenia patients’ learning difficulties and possibly certain clinical features, such as anhedonia (Horan, Blanchard, & Kring, 2006).

Several other factors could account for the patients’ poor performance on the PC task. One factor is the psychometric properties of the implicit learning tasks. Chapman and Chapman (1983, 1978) described the importance of developing cognitive tasks that are matched for reliability, difficulty, and variance levels when evaluating possible differential ability deficits in clinical populations. The implicit learning tasks selected for use in this study were not specifically designed to demonstrate a differential deficit, and they differed on these characteristics. The PC task demonstrated a split-half reliability coefficient of .81 and a mean difficulty level (proportion correct) of .65 (SD = .14), whereas the AG task demonstrated a split-half reliability coefficient of .60 and a mean difficulty level of .58 (SD = .09). In addition, differences in the formats of the tasks make direct comparisons difficult. For example, the PC task is about three times as long as the AG task, possibly contributing to its better reliability. Another key difference is that learning occurs gradually throughout the testing phase of the PC task, whereas the AG task comprises separate training and testing phases. Given these task differences, we cannot definitively conclude from the present study that patients showed a differential deficit on the PC task.

Another factor to consider is potential medication effects. Although impairment on the PC task in the schizophrenia group was not directly attributable to patients who were taking conventional antipsychotics, the atypical medications that most patients were taking do, to various extents, affect striatal dopamine transmission (Abi-Dargham & Laurelle, 2005), and this could have affected their performance (Beninger et al., 2003). Alternative research strategies, such as studying unmedicated and recent-onset patients or studying patients who were randomly assigned to different types of medications, could help to address this possible explanation.

Alternatively, as suggested by Weickert and colleagues (2002), the patients’ PC impairment could reflect disturbances in other aspects of neurocognition, rather than deficient cognitive implicit learning per se. For example, it has been proposed that the medial temporal lobe and associated explicit learning functions may interact (or compete) with the basal ganglia during the early stages of performing this task (Poldrack & Rodriguez, 2004). Furthermore, performance can be improved on this task even in patients with striatal pathology by recruiting medial temporal lobe regions (Moody, Bookheimer, Vanek, & Knowlton, 2004). Indeed, college undergraduates appear to be able to rely on medial temporal lobe regions in this task unless they are learning under cognitive load (Foerde et al., 2006). Although this study did not include a comprehensive neuropsychological assessment battery, the lack of correlation between the PC task and HVLT performance observed here argues against a contribution of explicit processing deficits to poor performance on the PC task in this study. Along these lines, it may be informative for later studies to conduct a fine-grained analysis of the characteristic learning strategy that individuals with schizophrenia use to perform the PC task, as has been done in studies of amnesic patients (Gluck, Shohamy, & Myers, 2002; Meeter, Myers, Shohamy, Hopkins, & Gluck, 2006).

The divergent pattern of implicit learning shown by individuals with schizophrenia may be useful for developing theoretical models of the mechanisms through which neurocognitive deficits affect daily life functioning and for guiding clinical interventions. Many social–cognitive processes that are required for adaptive social functioning appear to be learned and enacted automatically (Bargh & Williams, 2006; Lieberman, 2000; Satpute & Lieberman, 2006). As noted by Danion and colleagues (2001), intact aspects of implicit learning during development and everyday life may help explain why, despite pervasive explicit processing deficits, patients’ functional capacities are generally not grossly impaired. However, when automatic processing is guided by emotional contingencies, as is often the case in the social environment, disturbed aspects of implicit learning could contribute to patients’ social functioning difficulties (Bargh & Williams, 2006). In future research, it will be useful to determine whether performance on implicit learning tasks correlate with measures of social cognition or actual community functioning in schizophrenia.

In terms of clinical implications, schizophrenia patients’ ability to implicitly learn through repeated exposure without trial-by-trial feedback on the AG task appears to be a relative strength on which psychosocial rehabilitation can capitalize. For example, compensatory training techniques such as errorless learning, which minimizes the use of feedback to correct past mistakes and guide subsequent behavior, have proven useful in helping schizophrenia patients develop specific vocational skills and even social problem-solving skills (Kern et al., 2005; Kern, Liberman, Kopelowicz, Mintz, & Green, 2002). It remains to be determined whether the type of learning difficulty shown by patients on the PC task is amenable to restorative rehabilitation approaches.

References


IMPLICIT LEARNING IN SCHIZOPHRENIA


Received March 21, 2007
Revision received March 10, 2008
Accepted March 14, 2008