

An Event-Related fMRI Study of Artificial Grammar Learning
in a Balanced Chunk Strength Design

Matthew D. Lieberman

Grace Y. Chang

University of California, Los Angeles

Joan Chiao

Harvard University

Susan Y. Bookheimer

Barbara J. Knowlton

University of California, Los Angeles

Address Correspondence to:
Matthew D. Lieberman
Department of Psychology
Franz Hall
University of California
Los Angeles, CA 90095-1563
Phone: 310-206-4050
Fax: 310-206-5895
Email: lieber@ucla.edu

ABSTRACT

Artificial grammar learning (Reber, 1967) is a form of implicit learning in which cognitive, rather than motor, implicit learning has been found. After viewing a series of letter strings formed according to a finite state rule system, people are able to classify new letter strings as to whether or not they are formed according to these grammatical rules despite little conscious insight into the rule structure. Previous research has shown that these classification judgments are based on knowledge of abstract rules as well as superficial similarity (“chunk strength”) to training strings. Here we used event-related fMRI to identify neural regions involved in using both sources of information as test stimuli were designed to unconfound chunk strength from rule use. Using functional connectivity analyses, the extent to which the sources of information are complementary or competitive was also assessed. Activation in the right caudate was associated with rule adherence, whereas medial temporal lobe activations were associated with chunk strength. Additionally, functional connectivity analyses revealed caudate and medial temporal lobe activations to be strongly negatively correlated ($r=-.88$) with one another during the performance of this task.

Having good intuition about “what fits” and “what’s coming next” is essential to achieving one’s goals and meeting one’s obligations efficiently and effectively. Life is filled with scripts and recipes that have natural sequences and humans routinely take advantage of the predictability of these sequences to coordinate their thought and behavior (Lieberman, 2000; Schank & Abelson, 1977). Automating this knowledge of sequential regularities has the additional benefit that the relevant representations will be activated spontaneously in the presence of the sequential cues. Social interactions which depend on the simultaneous coordination of multiple processes undoubtedly benefit from an ability to automatically infer from facial expressions and tone of voice, a great deal of information about the intentions, evaluations, and personality (Ambady & Rosenthal, 1993; Chartrand & Jefferis, 2002; Lieberman & Rosenthal, 2001; Swann, Stein-Seroussi, & McNulty, 1992). Implicit learning refers to the ability to learn informational sequences when there is no conscious intent to extract this sequential information and no explicit knowledge that this information has been learned (Reber, 1993; Seger, 1994; Stadler & Frensch, 1998). Presumably, there are one or more neurocognitive mechanisms that are able to learn recognizable features and statistical relationships between cues in sequential patterns (Knowlton & Squire, 1996; Packard & Knowlton, 2002). However, there is a lack of consensus regarding the brain regions that support implicit learning, what these regions contribute computationally, and the extent to which the various neurocognitive mechanisms of implicit learning operate in a competitive or complementary fashion. There are at least three different reasons for this discord, each embedded in the methodologies used to study the neural substrates of implicit learning.

First, most of the existing studies have focused on neuropsychological populations. Several of these studies suggest that the basal ganglia are involved in implicit learning, while medial temporal areas may not be. Patients with Parkinson’s and Huntington’s disease have increasingly disturbed

basal ganglia function over time. Numerous studies have shown implicit learning to be absent or impaired in both groups (Martone et al., 1984; Heindel et al., 1989; Gabrieli, 1995; Ferraro et al., 1993; Knopman & Nissen, 1991; Knowlton, Mangels & Squire, 1996; Knowlton, Squire, et al., 1996), but spared in patients with medial temporal damage (Cohen & Squire, 1980; Knopman & Nissen, 1987; Knowlton, Mangels, & Squire, 1996; Knowlton, Squire et al., 1996). There have, however, been a number of studies showing that some forms of implicit learning are preserved in patients with Parkinson's or Huntington's disease (Daum et al., 1995; Gabrieli et al., 1995; Harrington et al., 1990; Reber & Squire, 1999; Smith et al., 2001; Peigneux et al., 1999). Some of the inherent limitations of neuropsychological investigations may prevent this research method from completely characterizing the neural bases of implicit learning. Parkinson's and Huntington's disease both have complicated progressions involving multiple neural structures over many years, and the rate of decline, both neurally and psychologically, show large individual differences (Grant & Adams, 1996). Of course, neuroimaging research has major limitations of its own, but the combination of neuropsychology with neuroimaging helps alleviate each method's weaknesses (Kosslyn & Intrilligator, 1992).

A second reason for the lack of consensus regarding the neural underpinnings of implicit learning follows from the recent suggestion that there are at least two different forms of implicit learning. Seger (1997) proposed that there are both motor and judgment-linked forms of implicit learning.¹ For instance, in the Serial Reaction Time task that has been used many times, the critical dependent measure is reaction time. In a typical instantiation of this task, a target appears in a sequence of locations on the screen and for each presentation, participants are required to press a button corresponding to the quadrant of the screen where the target appears as quickly as possible. Unbeknownst to participants, the sequence of presentation locations repeat every eight or so

presentations. Participants get faster over time as would be expected from basic practice effects. Nevertheless, when the underlying pattern is changed such that the old pattern is no longer present, reaction times increase, indicating a sensitivity to the underlying pattern despite little conscious knowledge thereof. With this implicit motor learning task, several neuropsychological studies suggest that good performance depends on intact basal ganglia (Ferraro et al., 1993; Jackson, et al., 1995; Knopman & Nissen, 1991; Willingham & Koroshetz, 1993). Along these same lines, five neuroimaging studies of the serial reaction time task have also implicated the basal ganglia (Berns, Cohen & Mintun, 1997; Grafton, Hazeltine, & Ivry, 1995; Hazeltine, Grafton, & Ivry, 1997; Rauch et al., 1995; Rauch et al., 1997).

Judgment forms of implicit learning have received much less attention in terms of their neural loci and thus it is unknown whether these forms of implicit learning rely on the same neural structures as motor forms of implicit learning. The oldest and most famous judgment implicit learning task is the artificial grammar task (Reber, 1967), in which letter strings are generated according to the rules of a Markovian grammar chain (see Figure 1). After being exposed to several “grammatical” exemplars that conform to the rules of the grammar chain, participants are informed of the grammar, without being shown the rules, and are asked to judge whether new letter strings conform to the grammar or not. Participants often express little confidence in their judgments, instead relying on their intuition or “gut feeling”. Nevertheless, unimpaired participants consistently perform above chance on this task. With neuropsychological tests of judgment implicit learning, the results have been equivocal at best. Both Huntington’s and Parkinson’s patients have shown preserved learning on the artificial grammar task (Peigneux, et al., 1999; Reber & Squire, 1999; Smith et al., 2001). However, on the weather task, another judgment implicit learning task, both groups were impaired (Knowlton et al, 1995; 1996), and two neuroimaging studies of the weather task reported basal ganglia activation

(Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Poldrack, Clark, Paré-Blagoev, Shohamy, Creso Moyano, Myers, & Gluck, 2001). Thus, it is unclear to what extent the basal ganglia are involved in judgment implicit learning despite their clear role in motor implicit learning.

The third concern focuses on an aspect of the methodology specific to artificial grammar research. The grammatical letter strings that are presented during the learning phase and the judgment phase of a given experiment tend to share two qualities that make them different from nongrammatical letter strings. First, by definition the grammatical items follow the *rules* while the nongrammatical items violate the rules at least once. The intention in most artificial grammar research has been to study how humans learn these implicit rules and the conditions under which such rule learning is enhanced or impaired. However, on average, grammatical test strings are also likely to bear greater *superficial similarity* to the learning strings than do the nongrammatical test strings. For the researcher this is an unfortunate confound, but for the participant this provides another route to perform well on the task. A number of researchers have demonstrated that the similarity of whole or partial test items to training items has an influence on grammaticality judgments (Perruchet & Pacteau, 1990; Vokey & Brooks, 1992; Meulemans & Van der Linden, 1997). For example, test items that contain letter bigrams and trigrams (2 and 3 letter chunks) that had been repeated frequently during training are more likely to be endorsed as grammatical than test items that do not contain frequently repeating bigrams and trigrams. Thus, sensitivity to bigram and trigram frequency could allow one to discriminate between grammatical and nongrammatical items in the absence of knowledge about rules.

Knowlton and Squire (1996) quantified the degree of superficial similarity between test items and learning items using a “chunk strength” metric. To compute chunk strength, all possible bigram and trigram “chunks” are identified for a test item. For instance, if a test item was XVT, there would

be total of three bigram and trigram chunks (XV, VT, XVT). An item's chunk strength, then, is the average number of times the chunks from a test item appear in any of the training items. Thus, test strings with a high chunk strength bear greater superficial resemblance to training strings. In most studies, the random creation of test items leads to grammatical items with higher average chunk strength than the nongrammatical items. In these cases, superficial similarity between test and training items may have been used to discriminate the grammatical and nongrammatical items to some extent, because grammaticality was confounded with chunk strength. One way to avoid this is to use a "balanced chunk strength design" in which the average chunk strength of grammatical items and the average chunk strength of nongrammatical items are equivalent. Although grammatical items naturally tend to have higher chunk strengths, it is possible to construct a high chunk strength nongrammatical item by using grammatical chunks at impermissible positions within the letter string. In a balanced chunk strength design, both high and low chunk strength grammatical and nongrammatical items are typically used in testing. Because the grammatical and nongrammatical items have the same average chunk strength, this design allows the experimenter to evaluate the effects of the grammatical rules independently of the effects of chunk strength. This design also enables the researcher to compare performance on high and low chunk strength items of the same grammatical status to evaluate subjects' sensitivity to chunk strength.

Knowlton & Squire (1996) used a balanced chunk strength design and found that both grammaticality and chunk strength make contributions to judgments of grammaticality. In this study, grammatical items of low chunk strength (GL) and nongrammatical items of high chunk strength (NGH) were included along with grammatical items of high chunk strength (GH) and nongrammatical items of low chunk strength (NGL). Individuals did rely on chunk strength when it was an available cue and thus endorsed NGH items to a greater extent than NGL items. In the

absence of chunk strength cues (i.e. low chunk strength items), people were able to discriminate test items based on grammaticality. These data suggest that chunk strength and grammaticality compete with one another such that chunk strength cues tend to override the use of grammaticality when both cues are present.

In a recent study, Chang and Knowlton (submitted) found further support for the idea that grammaticality and superficial similarity exert separate influences on judgments in this task. In one experiment, changing the font and case of letter strings between learning and test eliminated the influence of superficial similarity but not grammaticality. Changing these surface features left the grammatical structure of the cues intact while diminishing the superficial similarity between test and training items. In a second experiment, a dual-task procedure was also found to reduce accuracy related to chunk strength, but not grammaticality. This suggests that the processes associated with comparing test item chunks to training chunks are undermined by cognitive load, while the more implicit representations of the grammar continues to operate unimpaired.

There have been three previous neuroimaging studies of artificial grammar learning (Fletcher, Büchel, Josephs, Friston, & Dolan, 1999; Seger, Prabhakaran, Poldrack, & Gabrieli, 2000; Skosnik, Mirza, Gitelman, Parrish, Mesulam, & Reber, 2002). None of the studies employed a balanced chunk strength design. As such, these studies were able to identify regions associated with the general mode of making classification judgments in this task, but were not designed to identify regions specifically associated with rule adherence. Nor were they able to examine the extent to which rule use and chunk strength use rely on competitive or complementary neural processes.

In the current study, we have carried out an event-related functional magnetic resonance imaging (fMRI) study in which we used a balanced chunk strength design, independently manipulated grammaticality and chunk strength across test items. By manipulating both

grammaticality and chunk-strength, we were able to build on the previous imaging studies of artificial grammar learning (Seger et al., 2000; Skosnik et al., 2002) and assess the neural structures involved in these two dimensions of the artificial grammar task and the relationship between the two dimensions. Our primary hypothesis is that grammaticality and chunk strength knowledge will be associated with distinct neural substrates. For low chunk strength items, which are naturally less similar to training items, accurate discrimination of grammatical from nongrammatical items must rely on knowledge of the grammar rules. Because previous neuroimaging studies of implicit learning of rules and covariations have implicated the striatum (Poldrack et al., 2001; Seger & Cincotta, 2002), we predicted activation in the striatum associated with low chunk strength grammatical vs. nongrammatical items. We expected this activation to be present primarily in response to low chunk strength items because under these conditions, chunk strength cues that might lead to the use of different strategies are absent. We also expected to find changes in occipital areas replicating recent findings (Skosnik et al., 2002). Moreover, a growing literature suggests a competitive relationship between basal ganglia and medial temporal lobe processing (Packard, Hirsh, & White, 1989; Packard & Knowlton, 2002; Poldrack et al., 2001). Thus we expect that functional connectivity analyses will reveal a negative correlation between these regions in artificial grammar task performance.

RESULTS

Behavioral results

Of the nine participants, one participant's behavioral data was not collected because of a malfunction in the button box used to obtain responses. For the eight remaining participants, overall accuracy for the artificial grammar task was at 65.0%, which is significantly better than chance in a one-sample t-test, $t(7)=6.47$, $p<.001$. Grammatical items (GH and GL items) were endorsed more frequently than nongrammatical items (NGH and NGL items; Table 1). Participants endorsed 67.2%

of the grammatical items and 37.1% of the nongrammatical items, $t(7)=7.24$, $p<.001$. In addition, there was a tendency towards high chunk strength items (GH and NGH items) being endorsed more often than low chunk strength items. Participants endorsed 56.6% of high chunk strength items and 47.7% of low chunk strength items, $t(7)=2.33$, $p<.11$. The cell means, presented in Table 1, are all similar to the values obtained by Knowlton and Squire (1996) using an almost identical methodology.

Though we were unable to conduct a two-way analysis of variance (ANOVA) with so few participants, we were able to model the interaction term using a within-participants linear contrast (Rosenthal, Rosnow, & Rubin, 2000) with appropriate weights (-1, +1, +1, -1 for GH, GL, NGH, NGL, respectively). This contrast suggests the impact of item grammaticality was greater for low chunk strength items than for high chunk strength items, $t(21)=2.30$, $p<.02$, suggesting that in head to head competition chunk strength cues win over grammaticality in terms of cue utilization. Pairwise comparisons indicate that GL items were endorsed more frequently than NGL items (GL vs. NGL), $t(7)=6.66$, $p<.001$. Weaker, but still significant, GH items were endorsed more frequently than NGH items (GH vs. NGH), $t(7)=2.71$, $p<.04$. While there was no detectable impact of chunk strength on the endorsement of grammatical items (GH vs. GL), $t(7)=0.54$, $p>.2$, there was a marginal impact of chunk strength on the endorsement of nongrammatical items (NGH vs. NGL) such that NGH items were endorsed more frequently than NGL, $t(7)=2.12$, $p<.08$. We did not ask participants to respond quickly because we did not want our effects driven by motor phenomena, so it is not clear that our reaction times results are particularly meaningful. Nevertheless, we ran t-tests comparing the reaction times across all relevant conditions (G vs NG, H vs L, GH vs GL, GH vs. NGH, GL vs NGL) and there were no significant differences (all t 's < 1.01, all p 's > .30).

Functional imaging results

In the basic comparison between all grammatical and nongrammatical items, without regard for chunk strength, our group analysis yielded significant activations in basal ganglia, medial temporal lobe, inferior frontal cortex and claustrum (see Table 2). Given the previous work suggesting both abstract rules and exemplars are used in artificial grammar experiments that do not control for chunk strength (Knowlton & Squire, 1996), it is not surprising that regions associated with both abstract rule learning (caudate, putamen) and exemplar use (medial temporal lobe) were found.

Sensitivity to Grammaticality. To test our more specific predictions, we performed a series of contrasts that separated grammaticality from chunk strength effects. In order to best determine the regions sensitive to grammaticality, we compared grammatical versus nongrammatical items of low chunk strength (GL-NGL). Recall that low chunk strength items can only be differentiated on the basis of the abstract rules of the artificial grammar as they bear low superficial similarity to the exemplars of the grammar presented during training. As reported earlier, participants were most sensitive to grammaticality for low chunk strength items, endorsing 42.2% more grammatical than nongrammatical items. Comparing activations in the two conditions (GL-NGL) revealed caudate and occipital activations (see Table 2 and Figure 2a). The caudate activation is consistent with findings from other implicit learning paradigms such as the serial reaction time task and the weather prediction task (Poldrack et al., 2001; Rauch et al., 1995). The medial occipital deactivation is puzzling given that the GL and NGL items are visually similar, nonetheless, a recent neuroimaging study of non-declarative learning suggests that this deactivation may reflect greater processing fluency (Reber, Stark, & Squire, 1998). It should also be noted that the caudate activation found in this statistical contrast is not contiguous with the caudate activity found in the previous contrast comparing all grammatical and nongrammatical items.

Based on previous findings showing negative correlations between basal ganglia and hippocampal regions during implicit learning, we undertook an analysis of functional connectivity between these two structures during performance of this task. This analysis showed that greater caudate activation in this analysis was associated with reduced hippocampal activation. We obtained the percent signal change for each participant in a spherical volume surrounding the voxel of greatest statistical significance in the caudate (14, 16, 0). The signal change data were then entered into a simple regression analysis across participants to test the hypothesis that the activity of medial temporal areas would be negatively correlated with the activity of the caudate within this contrast (GL-NGL). This analysis identified a cluster within the right hippocampus (24, -32, 0) whose activity revealed a strong negative correlation ($r = -.87$) with the caudate. Such correlational analyses cannot establish causality between activations, however it does add to the growing body of evidence suggesting a competitive relationship between the basal ganglia and the medial temporal lobe (Packard & Knowlton, 2002; Poldrack, et al., 2001).

Sensitivity to Chunk Strength. Previous fMRI studies of artificial grammar learning have not separated out the effects of chunk strength from the effects of grammaticality, thus leaving the possibility that participants in those studies were responding to the superficial similarity between test and training items, rather than grammaticality. The preceding analysis examined the neural structures that were differentially activated by grammatical items when superficial similarity was not a useful cue (GL-NGL). The current analysis instead directly examines which neural structures *are* sensitive to the superficial similarity between test and training items. Comparing high chunk strength items to low chunk strength items revealed hippocampal and putamen activations, as well as an occipital deactivation (see Table 2). The hippocampal activation suggests that participants were retrieving training items when presented with high chunk strength test items as compared with low chunk

strength test items. The putamen activation suggests basal ganglia involvement in chunk strength processing. It is not entirely clear what to make of this putamen activation as it is typically associated with motor processing, however, Seger and Cincotta (2002) also found putamen activation in a non-motor task. Additional research is needed to specify the contribution of the putamen to different forms of implicit learning.

Despite the presence of hippocampal activity during high chunk strength trials, the caudate, rather than the hippocampus, was more active for grammatical high chunk strength items relative to nongrammatical high chunk strength items (GH-NGH; see Table 2). Thus, hippocampal activity appears to be associated with sensitivity to chunk strength cues, rather than grammaticality, while the right caudate is associated with sensitivity to grammaticality at both levels of chunk strength.

Competing cues for high chunk strength grammatical items. High chunk grammatical items can be accurately classified using either the grammaticality or chunk strength cues. Consequently, it might be expected that if there is a competitive relationship between the functioning of the basal ganglia, in support of abstract rule use, and the medial temporal lobe, in support of chunk strength use, we should see both structures activated by these items (GH). Furthermore, we should expect negative correlations between these structures across participants such that participants showing greater medial temporal lobe activations for these items should show less basal ganglia activations and vice versa. As expected, there were both hippocampal and caudate activations when high chunk strength grammatical items were compared to low chunk strength grammatical items (GH-GL; see Table 2 and Figure 2b). It should be noted that unlike the previous analyses, the caudate activation here is in the left hemisphere. A functional connectivity analysis was performed using the percent signal change from each participant in the 5mm sphere surrounding the point of maximal activity in the hippocampus (28, -30, 2) as a regressor. This analysis revealed a strong negative correlation ($r=-$

.88) between hippocampal activity and a cluster within the right caudate (14, 23, 0). This connectivity analysis suggests that participants with greater hippocampal activations in the GH-GL comparison tended to have smaller activations in the caudate, again suggesting the possibility of a competitive relationship between the basal ganglia and medial temporal areas. Alternatively, in additional analysis of high chunk strength nongrammatical items compared to low chunk strength nongrammatical items (NGH-NGL; see Table 2) for which only chunk strength can serve as a relevant cue, we found bilateral activations in the medial temporal lobes and no activations in the caudate.

Grammaticality vs. chunk strength sensitivity. In previous analyses, the caudate was found to be significantly active in the contrast for grammaticality and hippocampus was found to be active in the contrast for chunk strength. The implication, supported by functional connectivity analyses as well, is that the caudate and hippocampus are each selectively involved in one aspect of artificial grammar learning more than the other. To further examine this claim, we searched for hippocampal activation in the grammaticality contrast and caudate activation in the chunk strength contrast at lower statistical thresholds and neither appeared at a level of $p < .05$ uncorrected. Direct comparisons for each brain structure across the two contrasts were not significant, perhaps due to low statistical power. However, a meta-analytic comparison of all four activations (Rosenthal, 1991, formula 4.26), indicated that hippocampal activation for chunk strength and caudate activation for grammaticality together were marginally more significant than hippocampal activation for grammaticality and caudate activation for chunk strength together ($p < .08$). This supports the hypothesis that these structures are differentially involved in the two aspects of grammar learning.

DISCUSSION

In the present study, several different brain regions were identified that may contribute to classification in the artificial grammar task. The caudate nucleus was activated more by grammatical items than by items nongrammatical items. Alternatively, the hippocampus was activated more by high chunk strength items than by low chunk strength items. These results support the idea that artificial grammar judgments are multiply determined. Moreover, there was evidence of strong negative functional connectivity between the caudate and hippocampal activations suggesting a competitive relationship between the two regions.

The fact that participants are able to discriminate grammatical from nongrammatical items, even when chunk strength is held constant, indicates that they were able to acquire, abstract, rule-based information. In the present study, the effect of grammatical status on judgments was clearest for low chunk strength items, which were not superficially similar to training items. GL items were associated with greater activation in the right caudate nucleus compared to NGL items. The caudate activation for grammatical items was also present for items with high chunk strength. The caudate nucleus has been implicated in other implicit learning tasks, and has been suggested to play a role in the implicit acquisition and expression of regularities about the environment. For example, caudate activation accompanies performance of the serial reaction time task, in which participants implicitly learn to respond in a sequence of locations (Berns, Mintun, & Cohen, 1997). Caudate activation is also present when participants perform the weather prediction task, in which probabilistic cue-outcome associations are learned implicitly (Poldrack et al., 2001). The present data suggest that the application of implicitly learned rules in the artificial grammar task may also involve the caudate nucleus.

In contrast, caudate activation was not associated with chunk strength differences. High chunk strength items were associated with less activation in the occipital gyrus (BA 18) than were low

chunk strength items. High chunk strength items contained bigrams and trigrams that had been presented frequently during training, and thus these items may have benefited from perceptual fluency during test. A similar phenomenon has been reported for perceptual priming tasks, in which previously presented stimuli are processed more rapidly and are associated with decreased activation in visual areas compared to new items (Buckner & Koutstaal, 1998; Koutstaal et al., 2001). A decrease in activation has also been observed in a dot pattern classification task in which the target category is defined by similarity to a prototype pattern (Reber, Stark, & Squire, 1998). In this study, subjects were exposed to exemplars generated from a prototype, and were later asked to classify new items as to whether they belonged in this category. The test items that had been generated from the prototype were more likely to be classified into the category than randomly generated items, and these items were also associated with decreased activation in BA 18 than were the random items. This finding is consistent with the idea that items generated from the prototype benefitted from perceptual fluency, and this fluency could contribute to the classification of these items. The current results suggest that a similar mechanism may operate for chunk strength sensitivity in artificial grammar learning. The early locus of the deactivation in the visual system is consistent with the finding that chunk strength sensitivity is font specific (Chang & Knowlton, submitted).

High chunk strength items were associated with increased activation in the hippocampus, and for nongrammatical high chunk strength items there were additional regions of activation in medial temporal cortical regions. Because these regions have been closely associated with declarative memory, this activation likely reflects that participants were engaging in explicit retrieval of training exemplars when similar items were presented. This differential medial temporal lobe activation was more widespread for items that did not adhere to grammatical rules (NGH vs. NGL), perhaps because of reduced competition from the caudate. Previous research has shown that amnesic patients

exhibit normal sensitivity to chunk strength (Knowlton & Squire, 1994, 1996). Thus, it seems unlikely that explicit memory for training exemplars is *necessary* for normal performance. However, subjects may use, or may attempt to use, explicit memory for training items during classification.

The activation of the caudate nucleus associated with rule learning stands in contrast to neuropsychological data demonstrating normal performance by patients with Huntington's disease or Parkinson's disease (Knowlton, Squire, Paulsen, Swerdlow, & Swenson, 1996; Peigneux et al, 1999; Reber & Squire, 1999; Smith, et al, 2001; Witt, Nuhman, & Deuschl, 2002). In some of these studies (Reber & Squire, 1999; Knowlton et al, 1996; Smith et al, 2001) rule adherence and chunk strength were confounded in the test items, leaving open the possibility that patients were able to achieve normal performance without applying abstract rules. In another study using balanced chunk materials, patients with Parkinson's disease were able to classify items according to grammaticality during the first presentation but not in a subsequent presentation, suggesting that their knowledge of the grammatical rules may have been weaker than in controls (Peigneux, et al, 1999). In a fourth study, patients with Parkinson's disease were able to transfer knowledge of an artificial grammar to test items formed using a different letter set (Reber & Squire, 1999). These data suggest that patients with Parkinson's disease are able to acquire abstract rules in this task. However, because the caudate nucleus is less affected than the putamen early in Parkinson's disease (Canavan, Passingham, Marsden et al., 1989; Nahmias, Garnett, Firnau & Lan, 1985), the locus identified in the present study may have been spared in these patients, allowing them to apply implicit abstract rules in the classification task. The striatum is clearly a heterogeneous region, with functional differences along both medial-lateral and dorsal-ventral dimensions (Devan, McDonald, & White, 1999; Ashby, Noble, Filoteo, Waldron, & Ell, 2003). Thus, it is possible that patients with Parkinson's disease would show deficits on some striatal dependent tasks but not others. In addition, there may be ways to achieve

above chance performance after a letter set transfer using an alternative strategy such as relying on analogies between study and test items (Brooks & Vokey, 1991). In the absence of abstract knowledge, patients with Parkinson's disease may use such a strategy. Some recent neuroimaging evidence suggests that patients with Parkinson's disease show activation in brain structures associated with explicit memory during performance of an implicit memory task (Moody, Vanek, Bookheimer, & Knowlton, submitted). An intriguing possibility is that the disruption of striatal function in Parkinson's disease results in different brain systems to become involved in tasks that typically depend on the striatum.

One interesting finding in the present study was the negative functional connectivity between caudate and hippocampal regions during classification performance. A similar finding reported during performance of the weather prediction task (Poldrack et al., 2001; see also Lieberman, Jarcho, & Satpute, 2002) has been interpreted as indicating a competitive relationship between medial temporal-dependent explicit memory and basal ganglia-dependent implicit memory. Indeed, for high chunk strength items, which appear to activate medial temporal lobe, there was a smaller effect of grammaticality on endorsement rates as indicated by a significant interaction between chunk strength and grammaticality. There is some evidence that explicit retrieval attempts do not improve classification in the artificial grammar task and may even be detrimental (Reber, 1976, cf. Dienes, Broadbent, & Berry, 1991). The negative correlation seen between the medial temporal lobe and basal ganglia suggests a possible neural mechanism for this behavioral incompatibility.

The striatum has been implicated in a range of memory and executive functions (See Saint-Cyr, 2003) for a review. According to one view, the head of the caudate plays a role in explicit rule learning and working memory, while the tail of the caudate is involved in the implicit integration of information during category learning (Ashby & Waldron, 1999). Neuroimaging evidence has

provided partial support for this idea in that the head of the caudate activation has been shown in both explicit and implicit rule learning (Seger & Cincotta, 2002). The present results showing activation in the head of the caudate during the application of implicitly learned rules are consistent with these previous results.

In summary, the use of event-related fMRI, coupled with the separation of items by adherence to grammatical rules and chunk strength allowed us to identify distinct neural systems contributing to classification judgments and the task-dependent interaction between these systems. Although the artificial grammar learning task is one of the most widely studied implicit learning tasks, judgments appear to be based on multiple cognitive processes. The present study indicates that activation in the caudate nucleus may play a role in applying the implicitly learned rules that contribute to performance, that activation in the hippocampus and medial temporal lobe may play a role in retrieval of chunks from training items, and that these processes may operate in a competitive, rather than a complementary, fashion.

METHODS

Participants

Nine right-handed individuals (4 men and 5 women) participated in this study. Participants ranged in age from 20 to 32. Written consent in accordance with UCLA's Institutional Review Board's approved procedures was provided by each subject.

Materials

Grammatical letter strings were generated from the finite-state Markovian rule system shown in Figure 1, and were identical to those used in Knowlton and Squire (1996).² Strings were formed by starting at the IN arrow and following the diagram along the arrows, adding a letter at each transition from one arrow to the next until the OUT arrow is reached. Sixteen grammatical and 16

nongrammatical test item strings matched for chunk strength were presented twice during the test. Chunk strength was calculated as follows: The frequency of each bigram and trigram across all the training items was first calculated. The chunk strength of each of the test items was calculated by averaging this frequency for each bigram and trigram that occurred in the test item (Knowlton & Squire, 1996). Half of the grammatical and nongrammatical items were designated as high chunk strength items (GH, NGH), while the other half of each set were designated as low chunk strength (GL, NGL).

Procedure

During the training phase, twenty-three grammatical letter strings were shown to subjects one at a time on 3" x 5" notecards. After each string was presented for 3 seconds, the card was removed and subjects were asked to reproduce the string. If the string was not accurately reproduced, the string was shown again and the process was repeated up to three times before moving onto the next letter string. The presentation of these 23 strings was then repeated a second time.

Subjects were then positioned in the scanner and given the following instructions:

The letter strings that you were shown in the first phase of this experiment were all constructed using a very complex set of rules. All of the letter strings followed those rules. We are now going to show you several more letter strings while you are in the scanner and for each, we want you to tell us whether the new strings fits the rules or not; whether it's legal or not. The rules are very complex, so you probably won't be able to figure them out explicitly. Instead, we would like you to go with your gut feeling, your hunch, what may even feel like a guess when you are indicating whether the string is legal or not. On each trial you will see a cross in the middle of the screen for one second followed by a new letter string for one second. After the string disappears, please indicate whether the string follows the rules or not (demonstrate with the buttons). There will then be a pause of 14 seconds while the scanner gets ready to take the next set of pictures. At this point the cross will appear again and the process will repeat.

Over the course of four functional scans, each participant completed 64 trials. There were 16 each of the four trial types (GH, NGH, GL, NGL). On each trial the participant judged whether the presented stimuli adhered to the grammatical rules used to construct the stimuli from the training phase.

Image Acquisition

Images were acquired using a GE 3.0T MRI scanner with an upgrade for echo-planar imaging (EPI; Advanced NMR Systems, Inc.). A 2D spin-echo image (TR = 4000 ms; TE = 40 ms; matrix size 256 by 256; 4-mm thick, 1-mm gap) was acquired in the sagittal plane to allow prescription of the slices to be obtained in the remaining sequences. This sequence also ensured the absence of structural abnormalities in the brain of the enrolled subjects. For each subject, a high resolution structural T2-weighted echo-planar imaging volume (spin-echo; TR = 4000 ms; TE 54 ms; matrix size 128x128; 26 axial slices; 3.125-mm in-plane resolution; 4-mm thick, skip 1-mm) was acquired coplanar with the functional scans. Four functional EPI scans (gradient-echo; TR = 2000 ms; TE = 45 ms; flip angle = 76; matrix size 64 by 64; 13 axial slices; 3.125-mm in-plane resolution 4-mm thick, 1-mm spacing) were acquired, each for a duration of 4 minutes and 36 seconds. Each functional scan was composed of 138 brain volumes. The first five volumes were not processed due to initial signal instability in the functional scan. The final five volumes were not processed for each scan as well. The remaining 128 volumes corresponded to 8 volumes for each of 16 experimental trials. Each trial began with 1s of fixation followed immediately by a 1s presentation of the letter string to be judged. The letter string presentation was followed by 14s of blank screen to allow the participant to respond and to allow the hemodynamic response to return to baseline between trials. The 16 trials for each scan were composed of equal numbers of each trial type (GH, NGH, GL, NGL).

Data Analysis

We analyzed the imaging data using statistical parametric mapping (SPM'99; Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK). Images for each subject were first realigned to each other to correct for head motion, then normalized into a standard stereotactic space as defined by the Montreal Neurological Institute (provided in SPM'99), and smoothed with a 8mm Gaussian kernel, full width at half maximum, to increase signal to noise ratio. Comparisons were calculated using an significance level (height threshold) of $p < .001$ uncorrected for multiple comparisons with an extent threshold of 5 voxels. Our design was event-related and consequently the design was modelled using a canonical haemodynamic response and its temporal derivative (Friston, Zarahn, Josephs, Hensen, & Dale, 1999). Planned comparisons were computed as contrasts for individual participants. The resulting contrast images were then used in random effects analyses at the group level.

References

- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. Journal of Personality and Social Psychology, *64*, 431-441.
- Ashby, F.G., Noble, S., Filoteo, J.V., Waldron, E.M., & Ell, S.W. (2003). Category learning deficits in Parkinson's disease. Neuropsychology, *17*, 115-124.
- Ashby, F.G., & Waldron, E.M. (1999). On the nature of implicit categorization. Psychonomic Bulletin & Review, *6*, 363-378.
- Berns, G. S., Cohen, J. D., & Mintun, M. A. (1997). Brain regions responsive to novelty in the absence of awareness. Science, *276*, 1272-1275.

Brooks, L., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) & Mathews et al. (1989). Journal of Experimental Psychology: General, *120*, 316-323.

Buckner, R.L. & Koutstaal, (1998). Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. Proceedings of the National Academy of Sciences, *95*, 891-198.

Canavan, A. G. M., Passingham, R. E., Marden, C. D., Quinn, N., Wyke, M., & Polkey, C. E. (1989a). The performance on learning tasks of patients in the early stages of PD. Neuropsychologia, *27*, 141-156.

Chang, G. Y., & Knowlton, B. J. Sensitivity to Surface Features of Training Exemplars in Artificial Grammar Learning. (manuscript submitted for publication)

Chartrand, T. L., & Jefferis, V. E. (in press). Consequences of automatic goal pursuit and the case of nonconscious mimicry. In J. P. Forgas, K. Williams, & W. von Hippel (Eds.), Social judgments: Explicit and implicit processes. New York: Cambridge University Press.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. Science, *210*, 207-210.

Daum, I., Schugens, M. M., Spieker, S., Poser, U., Schonle, P. W., & Birbaumer, N. (1995). Memory and skill acquisition in PD and frontal lobe dysfunction. Cortex, *31*, 413-432.

Devan, B.D., McDonald, R.J., & White, N.M. (1999). Effects of medial and lateral caudate-putamen lesions on place- and cue-guided behaviors in the water maze: Relation to thigmotaxis. Behavioural Brain Research, *100*, 5-14.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, *17*, 875-887.

Ferraro, F. R., Balota, D. A., & Connor, L. T. (1993). Implicit memory and the formation of new associations in nondemented Parkinson's disease individuals and individuals with senile dementia of the Alzheimer type: A serial reaction time (SRT) investigation. Brain and Cognition, *21*, 163-180.

Fletcher, P., Büchel, C., Josephs, O., Friston, K., & Dolan, R. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. Cerebral Cortex, *9*, 168-178.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. NeuroImage, *10*, 607-619.

Gabrieli, J. D. E. (1995). Contribution of the basal ganglia to skill learning and working memory in humans. In J. C. Houk & J. L. Davis & D. G. Beiser (Eds.), Models of information processing in the basal ganglia (pp. 277-294). Cambridge, MA: MIT Press.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1995). Functional mapping of sequence learning in normal humans. Journal of Cognitive Neuroscience, *7*, 497-510.

Grant, I., & Adams, K. M. (1996). Neuropsychological assessment of neuropsychiatric disorders. New York: Oxford University Press.

Harrington, D. L., Haaland, K. Y., Yeo, R. A., & Marder, E. (1990). Procedural memory in Parkinson's disease: Impaired motor but not visuo perceptual learning. Journal of Clinical and Experimental Neuropsychology, *12*, 323-339.

Hazeltine, E., Grafton, S. T., & Ivry, R. B. (1997). Attention and stimulus characteristics determine the locus of motor-sequence encoding: A PET study. Brain, *120*, 123-140.

Heindel, W. C., Salmon, D. P., Shultz, C. W., Walicke, P. A., & Butters, N. (1989). Neuropsychological evidence for multiple implicit memory systems: A comparison of Alzheimer's, Huntington's & Parkinson's Disease Patients. Journal of Neuroscience, *9*, 582-587.

Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard, C. (1995). Serial reaction time learning and Parkinson's disease: Evidence for a procedural learning deficit.

Neuropsychologia, *33*, 577-593.

Knopman, D., & Nissen, M. J. (1987). Implicit learning in patients with probable Alzheimer's disease. Neurology, *37*, 784-788.

Knopman, D., & Nissen, M. J. (1991). Procedural learning is impaired in HD: Evidence from the serial reaction task (SRT). Neuropsychologia, *29*, 245-254.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. Science(273).

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, *20*, 79-91.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar depends on implicit acquisition of both abstract and exemplar-specific information. Journal of Experimental Psychology: Learning, Memory, and Cognition, *22*, 169-181.

Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., & Swenson, M. (1996). Dissociations within nondeclarative memory in Huntington's disease. Neuropsychology, *10*, 538-548.

Kosslyn, S. M., & Intriligator, J. M. (1992). Is cognitive neuropsychology plausible? The perils of sitting on a one-legged stool. Journal of Cognitive Neuroscience, *4*, 96-106.

Koutstaal, W., Wagner, A.D., Rotte, M., Maril, A., Buckner, R.L., Schacter, D.L. (2001). Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. Neuropsychologia, *39*, 184-199.

Lieberman, M. D. (2000). Intuition: A social cognitive neuroscience approach. Psychological Bulletin, *126*, 109-137.

Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2002). Self-processing in the X- and C-systems: The neural bases of schematicity. Manuscript submitted for publication.

Lieberman, M. D., & Rosenthal, R. (2001). Why introverts can't always tell who likes them: Multitasking and nonverbal decoding. Journal of Personality & Social Psychology, 80(2), 294-310.

Martone, M., Butters, N., Payne, M., Becker, J., & Sax, D. S. (1984). Dissociations between skill learning and verbal recognition in amnesia and dementia. Archives of Neurology, 34, 965-970.

Meulemans, T. & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. Journal of Experimental Psychology: Learning, Memory, & Cognition, 23, 1007-1028.

Moody, T.D., Bookheimer, S.Y., Vanek, Z., & Knowlton, B.J. An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. (manuscript submitted for publication).

Nahmias, C., Garnett, E. S., Firnau, G., & Lan, A. (1985). Striatal dopamine distribution in parkinsonian patients during life. Journal of Neurological Science, 69, 233-240.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. Annual Review of Neuroscience, 25, 563-593.

Peigneux, P., Meulemans, T., Van Der Linden, M., Salmon, E., & Petit, H. (1999). Exploration of implicit artificial grammar learning in Parkinson's disease. Acta Neurologica Belgica, 99, 107-117.

Perruchet, P. & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or fragmentary knowledge. Journal of Experimental Psychology: General, 119, 264-275.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. Neuropsychology, 13, 564-574.

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. Nature, 414, 546-550.

Rauch, S. L., Savage, C. R., Brown, H. D., Curran, T., Alpert, N. M., Kendrick, A., Fischman, A. J., & Kosslyn, S. M. (1995). A PET investigation of implicit and explicit sequence learning.

Human Brain Mapping, *3*, 271-286.

Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., Bush, G., Breiter, H. C., & Rosen, B. R. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. Human Brain Mapping, *5*, 124-132.

Reber, A. S. (1967). Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior, *6*, 855-863.

Reber, A. S. (1976). Implicit learning of synthetic languages: the role of instructional set. Journal of Experimental Psychology: Human Learning & Memory, *2*, 88-94.

Reber, A. S. (1993). Implicit learning and tacit knowledge: An essay on the cognitive unconscious. New York: Oxford University Press.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. Learning and Memory, *5*, 420-428.

Rosenthal, R. (1990). Meta-analytic procedures for social research. Newbury Park, CA: Sage.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). Contrasts and effect sizes in behavioral research: A correlational approach. New York: Cambridge University Press.

Saint-Cyr, J.A. (2003). Fronto-striatal circuit functions. Context, sequence, and consequence. Journal of the International Neuropsychological Society, *9*, 103-127.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals, and understanding. Hillsdale, NJ: Erlbaum.

Seiger, C. A. (1994). Implicit learning. Psychological Bulletin, *115*, 163-196.

Seger, C. A. (1997). Two forms of sequential implicit learning. Consciousness and Cognition, *6*, 108-131.

Seger, C.A. & Cincotta, C.M. (2002). Striatal activity in concept learning. Cognitive, Affective, and Behavioral Neuroscience, *2*, 149-161.

Seger, C. A., Prabhakaran, V., Poldrack, R. A., & Gabrieli, J. D. E. (2000). Neural activity differs between explicit and implicit learning of artificial grammar strings: An fMRI study. Psychobiology, *28*, 283-292.

Skosnik, P. D., Mirza, F., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2002). Neural correlates of artificial grammar learning. Neuroimage, *17*, 1306-1314.

Smith, J., Siegert, R. J., & McDowall, J. (2001). Preserved implicit learning on both the serial reaction time task and artificial grammar in patients with Parkinson's disease. Brain & Cognition, *45*, 378-391.

Stadler, M. A., & Frensch, P. A. (1998). Handbook of implicit learning. London: Sage.

Swann, W. B., Stein-Seroussi, A., & McNulty, S. E. (1992). Outcasts in a white-lie society: The enigmatic worlds of people with negative self-conceptions. Journal of Personality and Social Psychology, *62*, 618-624.

Vokey, J.R. & Brooks, L.R. (1992). Salience of item knowledge in learning artificial grammars. Journal of Experimental Psychology: Learning, Memory, & Cognition, *18*, 328-344.

Willingham, D. B., & Koroshetz, W. J. (1993). Evidence for dissociable motor skills in Huntington's disease patients. Psychobiology, *21*, 173-182.

Witt, K., Nühsman, A., & Deuschl, G. (2002). Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. Neuropsychologia, *40*, 1534-1540.

ACKNOWLEDGEMENTS

This research was supported in part by a McDonnell-Pew grant (JSMF 99-25 CN-QUA.05) to Matthew Lieberman and a National Science Foundation career award (IBN-9985417) to Barbara Knowlton. Correspondence should be addressed to Matthew Lieberman, Department of Psychology, Franz Hall, University of California, Los Angeles, CA 90095-1563 or by email: lieber@ucla.edu.

Table 1. Percentage of items endorsed as grammatical by condition

	Chunk Strength		
	<u>High</u>	<u>Low</u>	<u>Average</u>
Grammatical	65.6	68.8	67.2
Nongrammatical	47.7	26.6	37.1
Average	56.6	47.7	

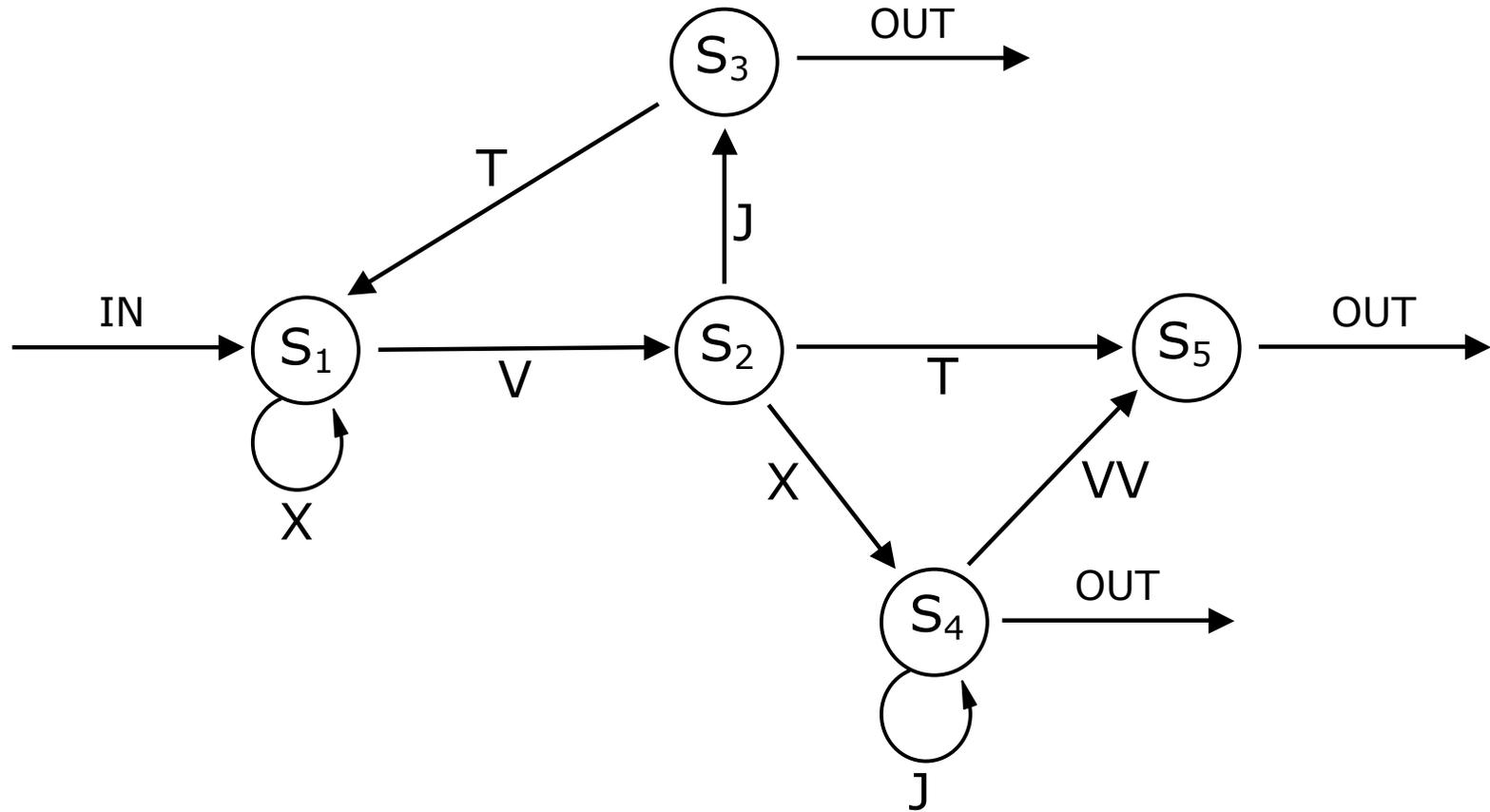
Table 2: Activations of all fMRI contrasts

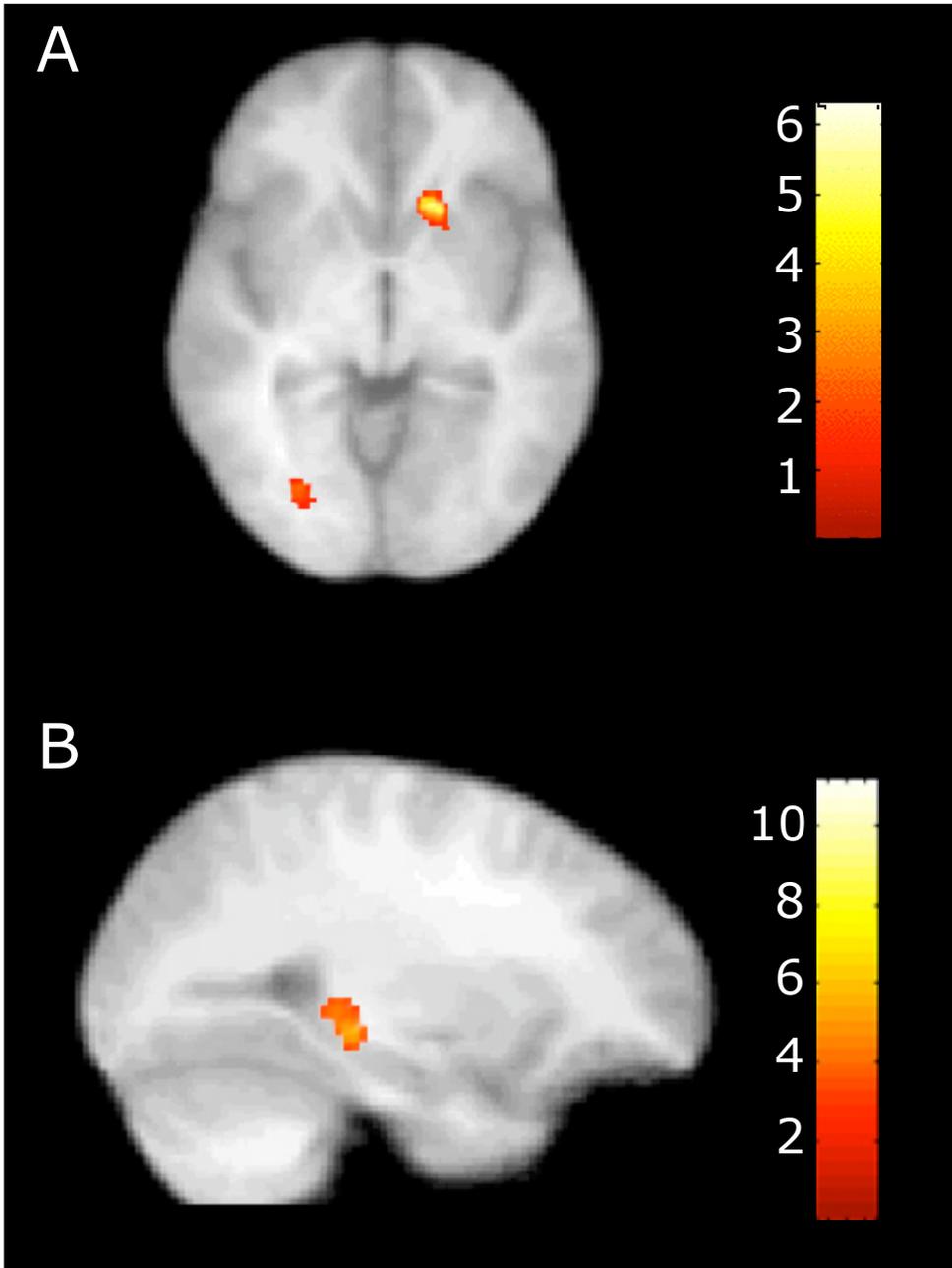
Region		Talarach Coordinate			Z-score
<i>grammatical minus nongrammatical – all items (G-NG)</i>					
Medial Temporal	L	-32	-42	-2	6.69
Caudate	R	12	-2	22	3.56
Putamen	L	-32	-10	20	4.89
Inferior Frontal	L	-44	10	18	4.11
Clastrum	R	28	14	10	4.91
<i>grammatical vs. nongrammatical – low chunk strength items (GL-NGL)</i>					
Caudate	R	14	16	0	5.28
Middle Occipital	L	-30	-74	-6	6.49
<i>high vs. low chunk strength – all items (GH+NGH)-(GL+NGL)</i>					
Hippocampus	L	-26	-34	-6	3.09
Putamen	L	-20	10	6	3.95
Occipital Gyrus	L	-16	-54	6	6 (deactivation)
<i>grammatical vs. nongrammatical – high chunk strength items (GH-NGH)</i>					
Caudate	R	17	-4	24	5.47
Caudate	R	34	-34	2	6.09
<i>high vs. low chunk strength – grammatical items (GH-GL)</i>					
Hippocampus	R	28	-30	-2	3.18
Caudate	L	-12	14	10	7.86
Superior Temporal	R	50	16	-6	6.31
Postcentral Gyrus	L	-58	-16	16	11.12
<i>high vs. low chunk strength – nongrammatical items (NGH-NGL)</i>					
Hippocampus	L	-26	-32	0	3.16
Medial Temporal	R	36	-36	2	4.70
Medial Temporal	L	-38	-18	-18	3.76
Middle Temporal	L	-48	-14	2	3.79

FIGURE CAPTIONS

Figure 1. The Markovian Grammar chain used to produce training and test items. Taken from Knowlton and Squire (1996). Strings are made by starting at IN and then following a path of arrows until OUT is reached. For each arrow traversed the indicated letter is added to the letter string.

Figure 2. Statistical activation maps in (a) Caudate activation (14, 16, 0) in the GL-NGL contrast and (b) Hippocampus activation (28, -30, -2) in GH-GL contrast. The color scale corresponds to Z-scores.





¹ By using the terminology of motor vs. judgment implicit learning, we do not mean to suggest the absence of judgment in motor forms of implicit learning. Rather, this designation is meant to highlight the presence of the motor response as a more integral component of this form of implicit learning as compared to judgment forms for which motor representations are less relevant.

² The grammar used in this study is less complex than other grammars that have been used in neuroimaging studies and therefore may be more amenable to explicit memory strategies. This grammar was used, however, because it has been used in multiple previous studies and is well characterized both behaviorally and neuropsychologically.