ELSEVIER

Neuropsychologia 42 (2004) 524–535

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Impaired probabilistic category learning in hypoxic subjects with hippocampal damage

Ramona O. Hopkins [a,b,*], Catherine E. Myers [c], Daphna Shohamy [d], Steven Grossman [d], Mark Gluck [d]

[a] Psychology Department and Neuroscience Center, Brigham Young University, Provo, UT, USA
[b] Deparment of Medicine, Pulmonary and Critical Care Divisions, LDS Hospital, Salt Lake City, UT, USA
[c] Department of Psychology, Rutgers University, Newark, NJ, USA
[d] Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ, USA

Received 2 December 2002; received in revised form 30 April 2003; accepted 4 September 2003

Abstract

Previous research has suggested that a probabilistic category learning task (e.g. weather prediction task) can be used to elucidate brain substrates of learning. We tested amnesic subjects with bilateral hippocampal damage due to hypoxia and matched controls on the weather prediction task and a variant, the "ice cream" task, which maintains a similar category structure. The hypoxic subjects were impaired relative to controls on both tasks; in the ice cream task, this difference was evident even early in training (first 50 trials). This finding is similar to functional neuroimaging (fMRI) studies in healthy subjects, which show medial temporal involvement even in early learning on this task. Additionally, strategy analysis of response patterns during learning suggest that the hypoxic group relied more heavily on simple, degraded learning strategies than did the control group. These results may suggest a qualification of the generally held conclusion that amnesic patients are not impaired at probabilistic category learning: at least under some circumstances, amnesic patients show an early and lasting deficit.

Keywords: Memory; Amnesia; Hypoxia; Category learning

1. Introduction

Recently, several studies have suggested that probabilistic categorization tasks can be used to examine the brain substrates of learning. One widely-studied categorization task, the "weather prediction" task (Knowlton, Squire, & Gluck, 1994), requires subjects to predict an outcome ("sun" or "rain") based on the appearance of cues ("tarot cards") which are each probabilistically associated with those outcomes (Fig. 1). It has generally been assumed that the probabilistic nature of the cue-outcome associations hinders direct memorization of the correct response, and that subjects instead gradually accrue information across trials (e.g., Knowlton et al., 1994). These findings have been taken to suggest that the brain systems involved in non-declarative memory might be more critical for this task than declarative memory systems, particularly as the task progresses.

Many behavioral studies have viewed the basal ganglia as supporting a non-declarative or "habit" learning system (Eichenbaum & Cohen, 2001; Robbins, 1996; Squire, 1994), and in fact individuals with Parkinson's disease, which disrupts basal ganglia function, are impaired on the weather prediction task (Knowlton, Mangels, & Squire, 1996). Conversely, individuals with anterograde amnesia, a specific deficit in acquisition of new declarative memories, performed similar to matched controls, at least on the first 50 training trials (Knowlton et al., 1994, 1996). Since the medial temporal lobes are often damaged in amnesic patients, these results suggest that the medial temporal lobes, although critical for many kinds of learning, may not be necessary for probabilistic category learning.

However, the interpretations of these prior results may be complicated by several factors. First, some of the amnesic patients tested by Knowlton et al. (1994, 1996) did not have medial temporal damage. Three of the eight amnesic subjects in the initial study had hippocampal damage confirmed by neuroimaging; two subjects had Korsakoff's disease (associated with diencephalic damage), one had medial thalamic

* Corresponding author. Tel.: +1-801-422-1170; fax: +1-801-422-0602.
E-mail address: mona_hopkins@byu.edu (R.O. Hopkins).

0028-3932/$ – see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2003.09.005

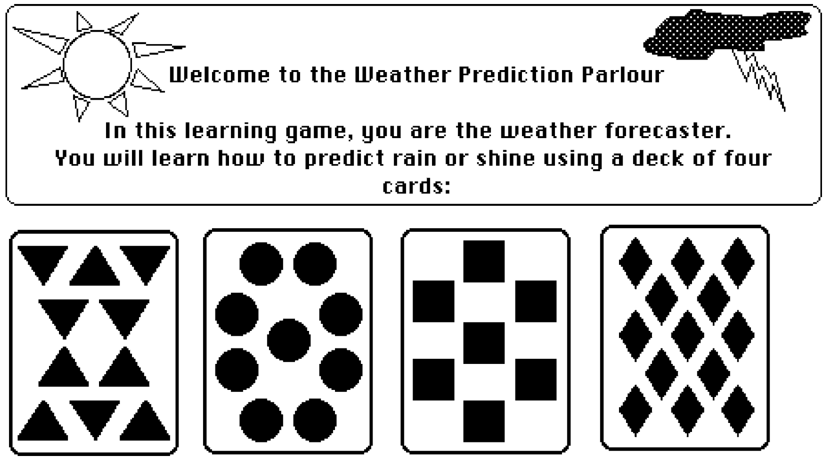

Fig. 1. The weather prediction task.

damage due to infarction, one had no lesions visible on neuroimaging, and one declined to undergo neuroimaging. Comparative studies suggest that the memory deficits may vary in patients with amnesia depending on the locus of brain damage (e.g., Leng & Parkin, 1989; Myers et al., 2001; Shoqeirat, Mayes, MacDonald, Meudell, & Pickering, 1990; Van der Linden, Bruyer, Roland, & Schils, 1993). Given these findings it is of interest to test individuals with particular amnesic etiologies on the weather prediction task.

Second, recent functional neuroimaging (fMRI) studies with healthy control subjects, using a slightly revised version of the weather prediction task, have demonstrated medial temporal engagement early in learning of the weather prediction task followed by deactivation as training progresses (Poldrack et al., 2001). These findings suggest that medial temporal structures are indeed engaged in this task, and that their involvement may occur even early in the training session. Such an interpretation is consistent with neurocomputational models of probabilistic category learning that suggest that the hippocampus and related medial temporal structures are important for setting up stimulus representations early in training. These representations may then be used by other brain regions such as the striatum, at which point, performance may be less dependent on the hippocampus (Gluck, Oliver, & Myers, 1996).

Hypoxic brain injury is an etiology that often causes fairly selective neuropathology in the hippocampus. For example, several recent studies have presented quantitative magnetic resonance imaging (MRI) analyses showing significant reduction in hippocampal volume in hypoxic subjects (e.g., Hopkins et al., 1995a; Hopkins, Kesner, & Goldstein, 1995b; Kesner & Hopkins, 2001; Press, Amaral, & Squire, 1989). Depending on the length and severity of the hypoxic episode, there may be non-specific degenerative neuropathology throughout the brain (Bachevalier & Meunier, 1996; Gale et al., 1999; Hopkins et al., 1995a). However, in select cases, cell death appears to be limited to (or particularly severe in) the hippocampus (e.g., Hopkins et al., 1995a,b; Kesner & Hopkins, 2001; Press et al., 1989; Zola-Morgan, Squire, & Amaral, 1986).

In this study, we tested nine individuals who survived hypoxic brain injury, with bilateral hippocampal damage verified by quantitative MR neuroimaging and anterograde amnesia verified by neuropsychological assessment, along with matched controls. In Task 1, we administered the weather prediction task, using the version that evoked medial temporal activation in the neuroimaging studies of Poldrack et al. (2001); this version differs from the original version used by Knowlton et al. (1994) only in that the stimulus-category mappings are slightly less random, meaning that subjects can guess the correct answer on up to 83% of trials, as opposed to 67% in the original version. This small change in paradigm appears to reduce subject frustration without causing any fundamental change in how healthy subjects approach the task or in overall performance (Gluck, Shohamy, & Myers, 2002; Poldrack et al., 2001; Shohamy, Onlaor, & Gluck, 2000; Shohamy, Onlaor, Myers, & Gluck, 2001). We expected that the revised version would similarly reduce subject frustration in the amnesic patients without fundamentally changing the probabilistic, incremental nature of the task.

In Task 2, we administered a second probabilistic category learning task, the "ice cream" task, which maintains the same logical structure as the weather prediction task, but uses a very different cover story and stimulus set (Shohamy et al., 2001). Specifically, whereas stimuli in the weather task were spatially discrete tarot cards, the stimuli in the ice cream task were toy faces (constructed from a Mr. Potatohead™ set) to be categorized based on various facial features. On each trial, the subject saw a potatohead figure and was asked to classify that figure according to whether it wanted chocolate or vanilla ice cream. As in the weather prediction task, there were four cues, each associated probabilistically with each outcome. If the hypoxic subjects show spared learning on these two tasks, this finding would support the claim that this general class of probabilistic category learning is independent of medial temporal lobe mediation, at least early in training.

In addition to assessing rate of learning, in terms of average correct responding across trials, we looked at individual subject data to assess what strategies subjects were using to approach this task. This strategy analysis was conducted by comparing a subject's response patterns against the pattern of responses that would be expected if the subject used a particular strategy. Probabilistic classification data from healthy young subjects suggests that normal subjects use a variety of strategies to approach this task (Gluck et al., 2002). If the hypoxic subjects show a similar pattern of strategy distribution, this would provide further support for the claim that this type of probabilistic classification learning is independent of medial temporal mediation. However, a different pattern of strategy distributions would suggest that hypoxic subjects use different strategies—and presumably different brain substrates—to approach the task, regardless of how well or poorly they perform overall.

2. Methods

2.1. Subjects

Nine individuals who became amnesic subsequent to hypoxic injury participated in the experiment. These individuals were recruited from LDS Hospital Rehabilitation Services and the Carbon Monoxide and Acute Respiratory Distress Syndrome (ARDS) studies. Table 1 shows demographic and etiology information for the hypoxic group. Seven hypoxic individuals (four female, three male) participated in Experiment 1; eight (five female, three male) participated in Experiment 2.

2.2. Neuroimaging

Hypoxic subjects underwent brain magnetic resonance (MR) imaging to verify hippocampal damage. MR images were acquired at 1.5 T with a quadriture head coil using standard clinical protocols. Sagittal T1-weighted (500/11/2; TR/TE/excitations) images were first acquired for localization, followed by coronal proton density and T2 weighted (3800/21; 105/2) spin echo images. Coronal interleaved sections were acquired with a section thickness of 3 mm. Coronal images were acquired on a 512×256 matrix with a 22 cm field of view.

Coronal intermediate and T2 weighted spin-echo images were obtained and processed using ANALYZE (Biomedical Imaging Resource, Mayo Foundation, Rochester, MN). The original 16-bit images are converted to 8-bit images in ANALYZE file format. Quantitative or volumetric analysis of the hippocampus and temporal lobes were carried out per the methods described previously (Bigler et al., 1997). Volumetric measures include the right and left hippocampal volumes and are reported as volumes (cm^3). Results were compared to age and gender matched control subjects from our normative imaging database (Bigler et al., 1997). Clinical brain MR reports by the neuroradiologist for the hypoxic subjects indicated no evidence of extra hippocampal lesions. No evidence of cerebellar or basal ganglia damage was found in the radiological reports or on neurological examination. Eight of the hypoxic subjects had bilateral hippocampal atrophy thought to be limited to the hippocampal region (quantified by magnetic resonance imaging). For one subject the hippocampus was not reduced in volume but focal lesions were present. See Fig. 2 for representative coronal scans through the hippocampus. The hypoxic subjects had significant hippocampal atrophy were compared to the normative control subjects ($t(6) = 4.4$, $P = 0.001$). There were no significant differences between hypoxic subjects and controls for temporal lobe volume ($P = 0.45$) or the perirhinal gyrus volume ($P = 0.39$) or parahippocampal gyrus volume ($P = 0.36$).

Although the hypoxic subjects have selective hippocampal damage, there may be more diffuse damage that may not be detectable using structural imaging. Cerebellar and basal ganglia damage appear unlikely, as confirmed by radiological reports and neurological examinations. Further, recent data suggest that basal ganglia lesions are less common

Table 1
Demographic and etiology information for the hypoxic group

ID	Sex	Age	Ed.	R Hipp	L Hipp	Etiology	Experiment 1	Experiment 2
32	Female	44	14	N/av.	N/av.	ARDS	Yes	Yes
33	Female	46	12	1.30	1.25	CRA	Yes	Yes
34	Male	35	12	1.26	1.31	CRA		Yes
35	Male	35	12	1.14	1.29	CRA	Yes	Yes
36	Male	34	14	Lesion	Lesion	CO poisoning	Yes	Yes
37	Female	47	12	2.20	2.30	Respiratory arrest	Yes	Yes
38	Male	44	12	1.26	1.31	CRA	Yes	
39	Female	38	12	N/av.	N/av.	ARDS	Yes	Yes
40	Female	49	12	N/av	N/av	CO poisoning		Yes
Mean		41.3	12.4	1.49	1.50		$N = 7$	$N = 8$
S.D.		5.8	0.9	0.42	0.41			

Age, Ed. (education) in years; R/L Hipp: right and left hippocampal volume in cubic centimeters (cm^3); N/av.: data not available. *Etiologies*: ARDS, acute respiratory distress syndrome; CRA, cardiac/respiratory distress; CO, carbon monoxide.

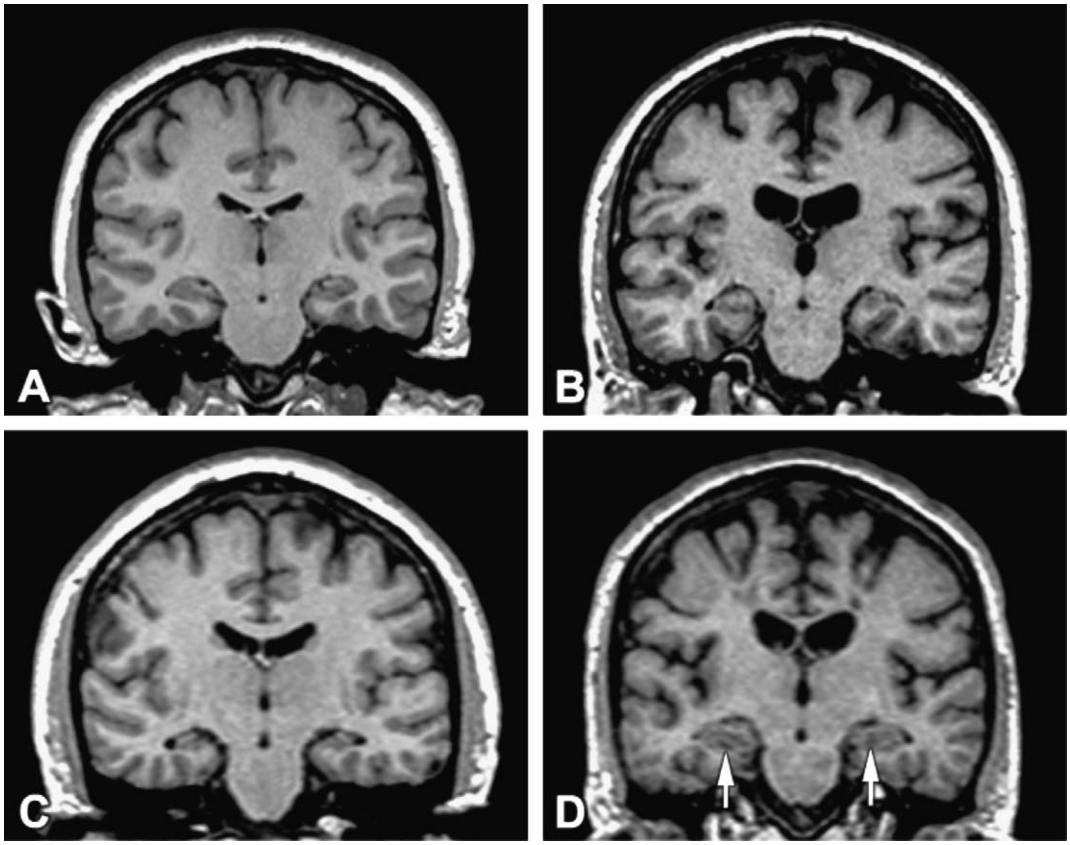

Fig. 2. Magnetic resonance images of coronal scans through the hippocampus and the level of the fourth ventricle. Panel A shows a normal control subject and panels B, C, and D show hypoxic subjects. Panels B and C show bilateral hippocampal atrophy and panel D shows hippocampal lesions in one hypoxic subject.

than previously reported following anoxia (Parkinson et al., 2002).

2.3. Neuropsychological tests

The hypoxic subjects were administered neuropsychological tests to assess general intellectual function and amnesic status. The tests include the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and Wechsler Memory Scale-Revised (WMS-R). The scores include the General Memory Index (GMI), a Delayed Recall Index, and an Attention/Concentration Index. These scores are age-adjusted with mean of 100 ± 15. To be considered amnesic, hypoxic subjects were required to score at least one S.D. below the mean on the Delayed Recall Index of the WMS-R, but within the normal range on the WAIS-R Full Scale Intelligence Quotient (FSIQ) and the WMS-R attention/concentration Index Score. Table 2 shows that all eight hypoxic subjects satisfied these criteria. Many hypoxic subjects received the minimum possible score of 50 on the WMS-R Delayed Recall Index. Alternatively the hypoxic subjects were within normal limits on the WMS-R Attention/Concentration Index and WAIS-R intelligence quotients, although group averages were slightly below the mean of 100, which would be expected for a group of unselected normal individuals. The hypoxic subjects received neuropsychological tests to assess verbal memory (logical memory subtest of the WMS-R) and the Rey Auditory Verbal Learning (RAVL) test (Table 3), visual memory (Rey-Osterrieth Complex Figure Test, ROCFT), and attention/concentration (digit subtest of the WAIS-R; Table 4).

Table 2
Neuropsychology test results for the hypoxic group

ID	GMI	Delay	WAIS-R			
			A/C	FSIQ	VIQ	PIQ
32	89	75	106	109	101	117
33	67	50	88	86	86	87
34	83	50	98	98	99	97
35	70	50	87	90	90	93
36	81	75	92	90	94	84
37	95	74	92	91	87	86
38	82	50	99	99	99	99
39	96	85	99	114	109	117
40	65	50	99	93	96	89
Mean	80.8	62.1	95.8	96.5	95.3	96.5
S.D.	11.5	14.7	6.3	10.0	7.7	13.5

WMS-R: Wechsler Memory Scale-Revised; GMI: General Memory Index; Delay: Delayed Recall Index; A/C: Attention/Concentration Index; WAIS-R: Wechsler Adult Intelligence Scale-Revised; FSIQ/VIQ/PIQ: Full-scale, verbal, and performance intelligence quotients.

Table 3
Rey Auditory Verbal Learning (RAVL) test results for hypoxic subjects (H1–H9) and control subjects (C1–C9)

	RAVL1	RAVL2	RAVL3	RAVL4	RAVL5	DELY	RECG
Hypoxics							
H1	5	6	7	8	8	5	7
H2	2	5	3	4	6	1	9
H3	4	7	6	8	8	0	3
H4	4	5	5	5	6	0	7
H5	6	8	10	10	9	6	14
H6	5	6	6	6	5	5	12
H7	4	7	6	8	8	3	8
H8	6	9	10	12	13	8	15
H9	2	2	3	3	4	0	0
Mean	4.2	6.1	6.2	7.1	7.4	3.1	8.3
S.D.	1.5	2.0	2.5	2.9	2.7	3.0	4.9
Controls							
C1	na	na	na	na	na	na	na
C2	na	na	na	na	na	na	na
C3	6	9	14	11	14	7	32
C4	9	10	13	15	15	15	25
C5	7	12	14	15	15	14	24
C6	9	11	12	13	14	11	24
C7	8	15	15	14	14	14	25
C8	7	12	12	12	14	9	24
C9	10	12	13	12	14	12	25
Mean	8	11.6	13.3	13.1	14.3	11.7	25.6
S.D.	1.4	1.9	1.1	1.6	0.5	2.9	2.9

Control subjects 1 and 2 did were not administered the test. Raw scores given for all measures and group scores presented as mean±S.D.. RAVL1, trial 1; RAVL2, trail 2; RAVL3, trial 3; RAVL4, trial 4; RAVL5, trial 5; DELY, delayed recall trial; RECG, recognition memory trial.

2.4. Control subjects

Nine healthy individuals were also recruited to serve as controls. To match the hypoxic group data, seven (five female, two male) controls participated in Task 1 and eight (five female, three male) participated in Task 2. Control subjects were screened for presence of any neurological or psychiatric condition (including depression) as well as for presence of any medication that could affect cognition. The healthy control group had a mean age of 39.1 years (S.D. 6.2) that was not significantly different from the hypoxic group (independent-samples t-test ($t(16) = 0.78$, $P = 0.443$), and mean education of 16.4 years (S.D. 2.2), which was higher than that of the hypoxic group ($t(16) = 5.09$, $P < 0.001$). The control subjects' VIQ was assessed using the North American Adult Reading Test (NAART; Blair & Spreen, 1989); this test generates a score, which can be converted to an estimate of VIQ with an expected mean of 100 (S.D. 15) in healthy normal subjects. VIQ in the control group was a mean 100.7 (S.D. 6.8) that did not differ significantly from the VIQ measure computed for the hypoxic subjects ($t(16) = 1.01$, $P = 0.329$).

The control subjects received neuropsychological tests to assess verbal memory (logical memory subtest of the WMS-R), visual memory (ROCFT), and attention/concentration (digit subtest of the WAIS-R; Table 3). Control subjects were recruited via the Memory Disorders Project at Rutgers-Newark, and received payment for participation at the rate of US$ 10 per hour.

All subjects signed statements of informed consent before initiation of behavioral testing. The study had IRB approval at LDS Hospital, Brigham-Young University, and Rutgers University, and conformed to institutional and federal guidelines for the protection of human subjects.

2.5. Apparatus

The experiment was conducted on a Macintosh i-book laptop computer, programmed in the SuperCard language (Allegiant Technologies, San Diego, CA). The keyboard was masked except for two keys, which the participant used to enter responses. For the weather prediction task, the keys were labeled "sun" and "rain". For the ice cream task, the keys were labeled "vanilla" and "chocolate". The experiment took place in a quiet testing room, with the participant seated at a comfortable viewing distance from the computer.

2.6. Task 1: weather prediction task

The weather prediction task was as previously described by Knowlton et al. (1994). Experiment 1, Task 2, using modified probabilities of Gluck et al. (2002), Shohamy et al. (2000) and Shohamy et al. (2001).

Briefly, the instructions stated that the participant was to learn to predict the weather ("sun" or "rain") based on four tarot cards (Fig. 1); on each trial between one and three cards were dealt, and the participant should enter a prediction by pressing the "sun" or "rain" key. After the participant responded, the correct answer was shown in the form of a sun or raincloud icon, which appeared above the stmuli. If the participant's response was correct, a score bar would increase and a smiley face would appear; otherwise, the score bar would decrease and a frowning face would appear. The stimuli remained on the screen until the subject responded or for 5 s maximum; if the subject did not respond within 2 s, a prompt appeared: "answer now".

Following Knowlton et al. (1994) and Gluck et al. (2002), all possible combinations of one, two or three cards were used. Each card was associated with each outcome with a fixed probability, as shown in Fig. 3: P("sun"|card 1) = 0.8, P("sun"|card 2) = 0.6, P("sun"|card 3) = 0.4, P("sun"|card 4) = 0.2. Thus, for example, card 1 was associated with sun on 80% of the trials on which it appeared and with rain on 20% of the trials on which it appeared. Similarly, cards 2, 3, and 4 were associated with sun on 60, 40, and 20% of trials, respectively. These probabilities were used to generate a series of 200 trials in which the outcomes sun and rain appeared equally often (Gluck et al., 2002). These trials were presented in a random but fixed order for all subjects. Cards could appear in any spatial order on the screen.

Table 4
Neuropsychology test results for hypoxic subjects (H1–H9) and control subjects (C1–C9)

	WAIS-R digit span	Logical memory immediate	Logical memory delay	ROCFT copy	ROCFT delay
Hypoxics					
H1	16	21	12	36	21
H2	17	1	0	32	5
H3	14	20	3	32	0
H4	8	14	3	32	5
H5	16	19	9	34	11
H6	12	25	17	34	12
H7	15	21	3	33	0
H8	17	26	20	36	18
H9	20	14	0	30	0
Mean	14.9	18.0	7.4	33.3	7.9
S.D.	3.4	7.6	7.4	1.9	7.9
Controls					
C1	15	36	35	36	22
C2	15	32	24	34	29.5
C3	15	32	31	35	22
C4	18	28	22	36	22
C5	15	32	32	36	20
C6	18	18	14	35	16
C7	14	26	26	36	29
C8	23	28	24	36	27
C9	24	33	29	36	27
Mean	17.4	29.4	26.3	35.6	23.8
S.D.	3.7	5.3	6.3	0.7	4.5

Percentile scores given for WAIS-R digit span and WMS-R logical memory immediate and delay recall. *Z*-scores given for ROCFT copy and delay recall portions based on norms published in Spreen and Strauss (1998, p. 352).

2.7. Task 2: ice cream task

Task 2 maintains the same logical structure as the weather prediction task from Task 1, but uses a different cover story and stimulus set (Shohamy et al., 2001). Specifically, whereas stimuli in the weather task were spatially discrete tarot cards, the stimuli in this experiment were toy faces (constructed from a Mr. Potatohead^{TM} set) to be categorized based on various facial features.

Stimuli were created using a Mr. Potatohead^{TM} set. All stimuli consisted of the basic Mr. Potatohead^{TM} face with black eyes, red nose, white arms and green feet. The face had a visible hole and "smiling" surface texture where the mouth would appear. This basic face was altered by addition of one or more facial features: cue #1: black hat, cue #2: black mustache, cue #3: red eyeglasses, cue #4: white bow tie. Fourteen stimuli were devised following the same logic as in the weather prediction task; thus, one, two, or three of the cues could appear on each trial. Fig. 3 shows some example stimuli. Each feature was associated with each outcome with the same probabilities as the four cards in the weather prediction task. Thus, the hat was associated with "vanilla" on 80% of the trials on which it appeared and with "chocolate" on 20% of the trials on which it appeared; the glasses, mustache, and tie were likewise associated with "vanilla" on 60, 40, and 20% of trials, respectively. These probabilities were used to generate a series of 200 trials in which the outcomes "vanilla" and "chocolate" appeared equally often. This trial order was identical to that used in the weather prediction task.

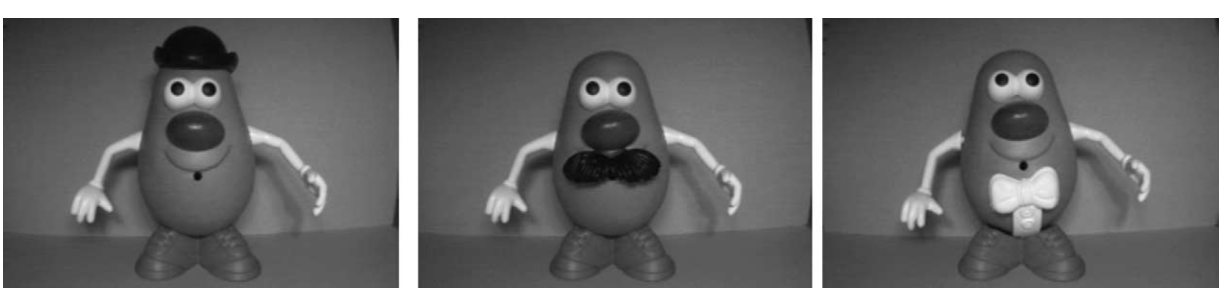

Fig. 3. Example stimuli in the ice cream experiment.

The subject was seated at a comfortable viewing distance from the computer screen. The following instructions were displayed:

> Welcome! In this game you are working in an ice cream shop. Customers will come in and buy vanilla or chocolate ice cream cones. Each time a customer visits, try to guess whether he wants vanilla or chocolate. If you guess correctly, you will earn an extra $1 tip. Try to collect as many tips as you can. Good luck!

On each trial, the screen showed a stimulus pattern (without ice cream) along with the prompt, *Which flavor do you think he wants*? (Fig. 4A). The subject responded by pressing one of the labeled keys; the word "vanilla" or "chocolate", corresponding to the subject's guess, appeared below the prompt. At this point, a picture of the same figure holding either a vanilla or chocolate ice cream cone (Fig. 4B) replaced the stimulus pattern. If the subject's guess was correct, the word "Correct" appeared at the bottom of the screen, a few coins were added to the image of the tip jar, and a sound of coins dropping was played through the computer speaker. If the guess was incorrect, the word "Incorrect" appeared at the bottom of the screen and the tip jar was unchanged. The stimulus (with ice cream) and feedback remained on the screen during a one-second inter-trial interval. The ice cream task was administered 6 months after the weather prediction task.

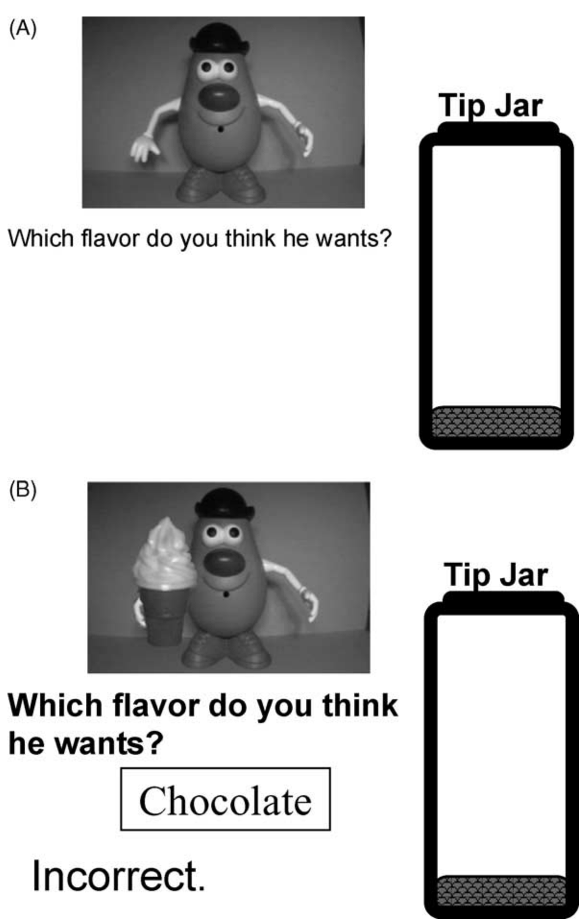

Fig. 4. Example screen events: (A) start of trial; (B) subject incorrectly chooses "chocolate".

2.8. Data analyses

Identical analyses were conducted for the both weather prediction and ice cream task data.

To facilitate comparison with earlier studies, we followed the analyses conducted in earlier studies (Gluck et al., 2002; Knowlton et al., 1994, 1996; Poldrack et al., 2001). On each trial, the computer recorded the pattern, the participant's response, and the actual weather. The participant's response was defined as correct if the participant chose the outcome, which was most often associated with that pattern, regardless of the actual outcome. For example, in Experiment 1, since pattern $A = 0001$ is most often associated with "sun", the "sun" response is optimal (and therefore correct) for that pattern—even though on a few trials the actual outcome will be "rain". Similarly, in Experiment 2, since pattern $A = 0001$ is most often associated with "vanilla", the "vanilla" response is optimal, regardless of the actual outcome. Note that there is no optimal response defined for patterns $F = 0110$ and $I = 1001$, which are equally often associated with each outcome. Using this scoring method, in each task, a subject who gave the correct answer on each trial could theoretically obtain a score of 100% correct, even though that subject would only have accurately predicted the actual weather on 83% of trials.

Following prior studies (e.g., Knowlton et al., 1994, 1996), we analyzed learning over the entire experiment in blocks of 50 trials as well as over the first 50 trials in blocks of 10 trials.

2.9. Strategy analyses

To investigate how hypoxic and control subjects approached learning in the two category learning tasks, we conducted strategy analysis using the methods presented in Gluck et al. (2002). In that paper, response data from most subjects could be accurately described by one of several "strategy models". Each model was constructed by considering the ideal data that would be expected if a hypothetical subject were strictly following a particular response strategy. For example, the optimal strategy to approach this class of task is a *multi-cue* strategy, also known as probability maximizing, defined as responding to each pattern with the response that is most often correct for that pattern; a subject reliably using this strategy could obtain up to 100% optimal responding. A second strategy is a *one-cue* strategy, in which subjects base responding on the presence or absence of a single cue and ignore the other cues; for example, in the weather prediction task, a subject might respond "sun" whenever the square card is present and "rain" whenever it is absent. A subject reliably following a one-cue strategy based one of the two highly-predictive cues could obtain up to 90% optimal responding. A third strategy, the *singleton* strategy, involves learning the optimal response to each of the four patterns (A, B, D, H) where a single cue (card or facial feature) is present, and guessing on the remaining patterns where two or more features are present in conjunction. A subject reliably using this strategy could obtain up to 65% optimal responding. While these three strategies—probability maximizing, one-cue and singleton—clearly do not exhaust the list of possible strategies, they appear to account adequately for the vast majority of subject data (Gluck et al., 2002).

To determine which strategy model best described an individual subject's data in the current experiment, we compared the subject's data trial-by-trial against the ideal data expected if that subject were reliably following that strategy, and took the squared difference between the responses generated by the subject vs. the ideal, normalized by the number of patterns (see Gluck et al., 2002, for full mathematical description). The result was a score between 0 and 1 for each strategy, with 0 indicating a perfect fit between the ideal data and the actual subject data. Comparing against all strategies examined, the one generating the lowest score was designated the *best-fit* strategy model for that particular subject.

Because subjects might shift strategies during the course of learning in an attempt to improve performance, we calculated best-fit strategies for each subject across the last block (50 trials) of learning, when learning and strategy were most likely to have stabilized.

3. Results

3.1. Neuropsychological tests

Table 3 shows neuropsychological data for the control and hypoxic groups for the RAVL. A repeated-measures ANOVA with group as the independent variable, for trails 1–5 revealed a significant group effect ($F(1, 14) = 43.76$, $P < 0.001$) with significant effect of trial ($F(4, 56) = 44.02$, $P < 0.001$) and group-trial interaction ($F(4, 56) = 5.61$, $P = 0.001$). Multiple independent ANOVAs, with alpha adjusted to 0.01 to protect significance levels, revealed group differences on each of the five trials (all $P < 0.001$). There was a significant group effect of delayed recall ($F(1, 14) = 32.82$, $P < 0.001$), and recognition ($F(1, 14) = 67.77$, $P < 0.001$).

Table 4 shows neuropsychological data for the control and hypoxic groups for the digit span, logical memory, and ROCFT.

A repeated-measures ANOVA, with group as the independent variable, on these six measures revealed significant group effects ($F(1, 16) = 38.48$, $P < 0.001$), as well as within-subject effects ($F(5, 80) = 48.53$, $P < 0.001$) and group-test interactions ($F(5, 80) = 12.05$, $P < 0.001$). Multiple independent-samples *t*-tests, with alpha adjusted to 0.008 to protect significance levels, revealed no group difference on digit span or ROCFT copy (all $P > 0.01$), but significant differences on the immediate and delay portions of the logical memory test and on the immediate and delay recall portions of the ROCFT (all $P < 0.005$). Thus, the hypoxic group was selectively impaired on both immediate and delay memory measures, relative to the control group.

3.2. Task 1: weather prediction task

Over all 200 training trials, the control group averaged 83.8% correct responses (S.D. 8.3) while the hypoxic group averaged 61.6% correct responses (S.D. 10.6). Fig. 5A shows the percentage correct responding over all 200 trials in blocks of 50 trials. There was a significant group effect ($F(1, 11) = 11.61$, $P = 0.006$), with no effect of education level ($F(1, 11) = 0.07$, $P > 0.500$), no within-subjects effect of block ($F(3, 33) = 0.38$, $P > 0.500$), and no interactions (all $P > 0.500$). Fig. 5B shows the percentage correct responding over the first 50 trials. A repeated-measures ANOVA found that the group effect fell short of statistical

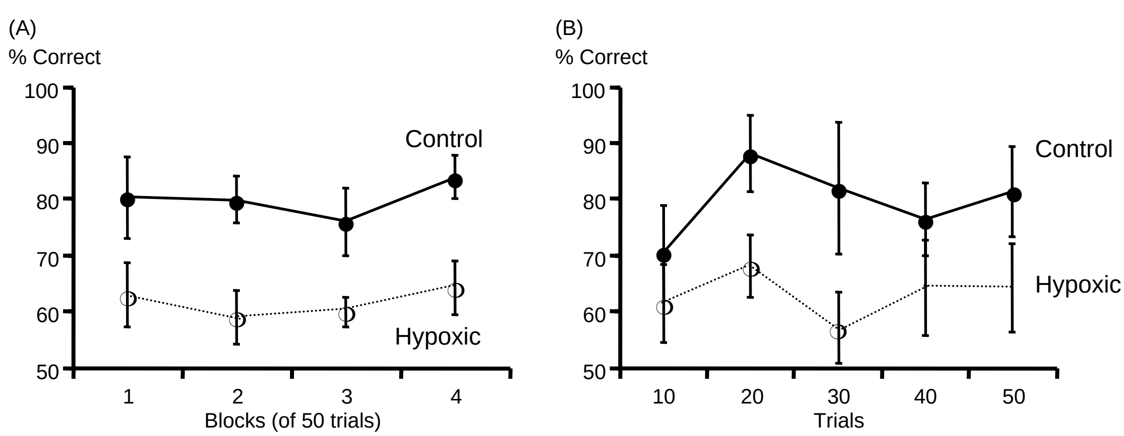

Fig. 5. Experiment 1 data: (A) learning curve over all 200 trials; (B) learning curve over first 50 trials.

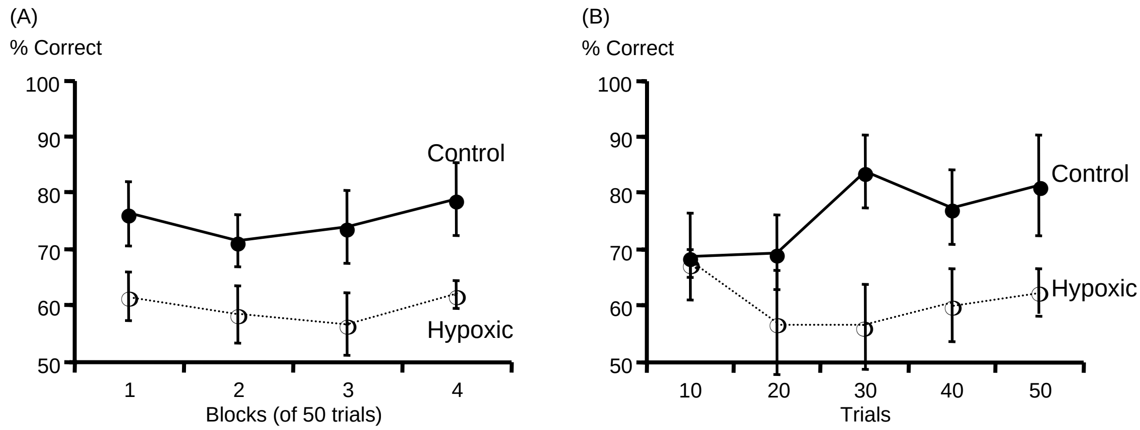

Fig. 6. Experiment 2 data: (A) learning curve over all 200 trials; (B) learning curve over first 50 trials.

significance ($F(1, 11) = 3.63$, $P = 0.083$), with no effect of education level ($F(1, 11) = 0.46$, $P > 0.500$), no within-subjects effect of trial ($F(4, 44) = 1.59$, $P = 0.195$), and no interactions (all $P > 0.500$). However, this lack of group difference was due to one control subject who achieved less than 50% correct in block 1 (this subject later went on to achieve over 70% correct in subsequent blocks). If this subject's data were not included, the difference between groups is evident even in the first 50 trials (repeated-measures ANOVA; group: $F(1, 10) = 7.87$, $P = 0.019$); education: $F(1, 10) = 0.18$, $P > 0.500$); trial: $F(4, 40) = 1.78$, $P = 0.152$; interactions: all $P > 0.100$).

Although the effects of block fell short of statistical significance, there was a trend for learning in the control group: neither group was responding significantly above chance during trials 1–10 (controls: mean 67.8, S.D. 23.5; hypoxics: mean 61.4, S.D. 18.6; all $P < 0.05$); by the end of the experiment (trials 191–200), the controls but not hypoxics were responding significantly above chance (controls: mean 86.6, S.D. 10.0, $P < 0.05$; hypoxics: mean 76.2, S.D. 17.5, $0.05 < P < 0.10$).

Using a performance criterion of 65% correct responses (Poldrack et al., 2001), all seven controls but only three hypoxic subjects were performing at or above this level by the final block of 50 trials.

3.3. Task 2: ice cream task

Over all 200 trials, the control group averaged 76.8% correct responses (S.D. 14.7) while the hypoxic group averaged 59.8% correct responses (S.D. 10.2). Fig. 6A shows the percentage correct responding over all 200 trials in blocks of 50 trials. There was a significant group effect ($F(1, 13) = 7.68$, $P = 0.016$), with no effect of education level ($F(1, 13) = 0.92$, $P = 0.355$), no within-subjects effect of block ($F(3, 39) = 1.49$, $P = 0.233$), and no interactions (all $P > 0.100$). Fig. 6B shows the percentage correct responding over the first 50 trials. A repeated-measures ANOVA found a significant effect of group ($F(1, 13) = 5.21$, $P = 0.04$), with no effect of education level ($F(1, 13) = 0.19$, $P > 0.500$), no within-subjects effect of trial ($F(4, 52) = 1.36$, $P = 0.260$), and no interactions (all $P > 0.100$).

As in the weather task, neither group was responding significantly above chance during trials 1–10 (controls: mean 70.0, S.D. 20.0; hypoxics: mean 63.8, S.D. 7.4; all $P > 0.05$); by the end of the experiment (trials 191–200), both groups were responding significantly above chance (controls: 78.1, S.D. 11.9; hypoxics: 75.0, S.D. 9.9; all $P < 0.05$).

Using a performance criterion of 65% correct responses (Poldrack et al., 2001), six of the eight controls but only three hypoxic subjects were performing at or above this level by the final block of 50 trials.

3.4. Strategy analysis

Fig. 7A and B shows the results of the strategy analysis for both the weather prediction and ice cream tasks. Across both tasks, all but one of the control subjects were well-fit by one of the three strategies investigated; there was a distribution across these strategies. By contrast, although most hypoxic subjects were also well-fit by one of the strategies, most were best-fit by a singleton strategy. There was no obvious relationship across tasks among subjects who were best-fit by various strategies: although three of the hypoxic subjects were best-fit by singleton in both tasks, the remaining three were best fit by different strategies in the two tasks; only one control was best-fit by the same strategy (a one-cue strategy) in both tasks. Collapsing data across the two experiments, there was a highly-significant difference in strategy distribution between hypoxic and control subjects (χ^2-test, $\chi^2(3) = 14.53$, $P = 0.002$), with hypoxic subjects appearing to rely more heavily on the singleton strategy than controls.

Fig. 7C shows the relationship between best-fit strategy and performance, collapsed across the two tasks. Although in general subjects best-fit by a multi-cue or one-cue strategy outperformed those subjects best-fit by a singleton or other

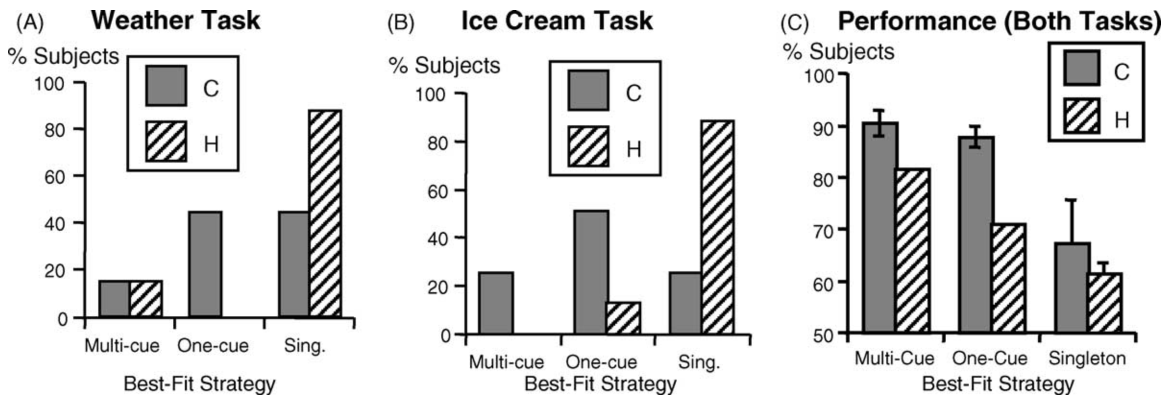

Fig. 7. Results of strategy analysis in control (C) and hypoxic (H) subjects. (A and B) Percentage of subjects' best-fit by multi-cue, one-cue and singleton models in the (A) weather task and (B) ice cream task. (C) Relationship between performance and best-fit strategy; vertical bars are standard error for those groups with $n > 1$.

strategy, these differences fell short of statistical significance (ANOVA, block 4 performance × best-fit strategy; controls: $F(3, 11) = 1.64$, $P = 0.237$; hypoxics: $F(2, 12) = 2.38$, $P = 0.134$).

4. Discussion

Consistent with earlier findings that amnesia impairs probabilistic category learning (Knowlton et al., 1994), we found that individuals with MT amnesia resulting from hypoxic brain injury were impaired both on the weather prediction task and on the ice cream task. Knowlton et al. found that the amnesic deficit emerged late in training, but that amnesics were not impaired early in training (e.g. across the first 50 trials). In our study, the difference between amnesic and control groups was apparent even within the first 50 trials in the ice cream task, while the group difference in performance on the weather task visible in Fig. 4B failed to reach significance only because of one control subject who performed very poorly on this block. In both tasks, there was no explicit effect of block; however, controls did appear to show some learning, since in both tasks, control scores did not differ from chance on the first 10 trials, but were significantly above chance by the last 10 trials.

In addition to slower learning, the hypoxic group appeared to be using qualitatively different strategies to approach learning. Fig. 6 shows that while controls tended to use a variety of strategies, including the optimal multi-cue strategy, hypoxic subject data were overwhelmingly consistent with a degraded singleton strategy, in which subjects learn the correct answer to a few "easy" patterns, and guess on the rest. Interestingly, a prior study suggested that healthy subjects tend to start out with simple strategies, such as the singleton strategy, but gradually shift to more effective strategies, like the multi-cue strategy, with extended training (e.g. 600 trials; Shohamy et al., 2000). It is possible that hypoxic subjects in the present study were merely slowed at learning, but would have shifted to multi-cue strategies with extended training. Alternately, it is possible that the hypoxic group was qualitatively unable to adopt multi-cue strategies; this would be generally consistent with the idea, prevalent in the animal literature, that the hippocampus and associated medial temporal structures, are important for configural or relational learning but not for learning simple cue-outcome associations (e.g. Sutherland & Rudy, 1989; Eichenbaum, Otto, & Cohen, 1994). The present data do not distinguish these possibilities, but extended multi-day training might help establish whether hypoxics are transiently or qualitatively impaired at these tasks.

In either case, our finding of impaired probabilistic category learning in amnesic subjects is similar to, but slightly different than, the prior data of Knowlton et al. (1994), who found that amnesic subjects differed from controls late, but not early, in training.

One factor that may contribute to the difference between studies is obviously the difference in subject group. In the current study, the amnesic subjects sustained bilateral hippocampal damage as a result of hypoxic brain injury. In the original Knowlton et al. (1994) study, the amnesic subjects had etiologies including hypoxia, diencephalic damage, and one subject who became amnesic with no known precipitating event. It is possible that these differences in etiology could be relevant. It should be noted that individuals who become amnesic following hypoxic brain injury may have extra hippocampal damage, and this is often reflected on deficits in cognitive and attentional, as well as memory, tests. While the neurologic exam, clinical and quantitative MR analysis of the hypoxic subjects in the current study did not find evidence for extra hippocampal damage it is possible that structural changes not observed on neuroimaging or functional changes that may be found using functional imaging techniques. Similarly, while the neuropsychological results summarized in Tables 2 and 3 for cognitive (VIQ, etc.) and attentional (digit span) measures were in the average to above average range and not significantly different from controls, the scores are numerically lower than the VIQ and digit span scores obtained in our control group. Thus, it is possible that the hypoxic subjects' scores may contribute to the deficits observed on this type of probabilistic category learning task. However given that the observed absence of extra hippocampal damage, scores on VIQ and digit span in the average to above average range, and the fact that hypoxic subjects did not differ from the controls on these tests, the observed differences in performance on the probabilistic category learning tests are not likely due to neuropsychological test performance.

A second difference between our study and the prior Knowlton et al. (1994) study is that in the prior study mean subject age was over 63 years, which is considerably older than the subjects in the current study, whose mean age was approximately 41 years. Pilot studies in our laboratory have suggested that there may be significant age effects on this task, with healthy older subjects performing worse than younger subjects on average (Abu-Shaba, Myers, Shohamy, & Gluck, 2001). It is possible that age contributed to the impairments in Knowlton's control subjects, masking differences between controls and amnesic subjects that would have emerged with younger subjects.

It is also the case that our weather prediction task differed from that of Knowlton et al. (1994; Experiment 1, Task 2) in the specific cue-outcome probabilities. Knowlton et al. used probabilities of 75, 57, 43, and 25% for the four cards; following more recent studies (Gluck et al., 2002; Poldrack et al., 2001; Shohamy et al., 2000, 2001), we used the cue-outcome probabilities of 80, 60, 40, 20%. In other words, the categories in our task were slightly more discriminable, which may make the task somewhat easier to master. However, in terms of the actual frequency with which each pattern was associated with each outcome, the difference between versions is quite small. For example, pattern A (only cue #1 present) is associated with the sun/vanilla outcome on 89% of trials in our version and on 85% of trials in the Knowlton version. If it is indeed the increased differentiability in our version of the task which accounts for the variation between our results and Knowlton et al.'s results, this suggests that probabilistic category learning in amnesia is exquisitely sensitive to very small variations in the cue probability structure, and that changing the outcome on a small proportion of trials has a large effect in which brain structures are, or are not, critical in early learning.

In either case, the original conclusion that probabilistic category learning is not disrupted in amnesia seems to require qualification. In our study, individuals with amnesia due to hypoxic brain injury showed an early and lasting impairment on two probabilistic category learning tasks. These amnesic individuals had documented bilateral hippocampal atrophy, and while the presence of extra hippocampal damage cannot be completely ruled out, it certainly appears possible that medial temporal damage affects probabilistic category learning, at least under some procedural circumstances, and in some amnesic etiologies.

Future work will be required to determine whether it is indeed the case that different amnesic subgroups, with different etiologies and patterns of brain damage, show different levels of impairment on probabilistic category learning tasks, or whether it is merely the case that amnesic patients in general are spared in the early stages of "hard" category learning tasks (as in the prior Knowlton et al., 1994 study), but impaired relative to controls on more easily discriminable versions.

Acknowledgements

This research was supported by a grant from the National Institute of Mental Health (R01 MH065406-01A1). We would like to thank Craig Vickers for his assistance with this manuscript.

References

Abu-Shaba, N., Myers, C., Shohamy, D., & Gluck, M. (2001). Age effects in probabilistic category learning. Newark: Rutgers University.

Bachevalier, J., & Meunier, M. (1996). Cerebral ischemia: Are the memory deficits associated with hippocampal cell loss? *Hippocampus*, *6*, 553–560.

Bigler, E. D., Blatter, D. D., Anderson, C. V., Johnson, S. C., Gale, S. D., Hopkins, R. O., & Burnett, B. (1997). Hippocampal volume in normal aging and traumatic brain injury. *American Journal of Neuroradiology*, *18*, 11–23.

Blair, J., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, *3*, 129–136.

Eichenbaum, H., & Cohen, N. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press.

Eichenbaum, H., Otto, T., & Cohen, N. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, *17*, 449–518.

Gale, S. D., Hopkins, R. O., Weaver, L. K., Bigler, E. D., Booth, E., & Blatter, D. D. (1999). MRI, quantitative MRI, SPECT, and neuropsychological findings following carbon monoxide poisoning. *Brain Injury*, *13*, 229–243.

Gluck, M., Oliver, L., & Myers, C. (1996). Late-training amnesic deficits in probabilistic category learning: A neurocomputational analysis. *Learning and Memory*, *3*, 326–340.

Gluck, M., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task? Individual variability in strategies for probabilistic category learning, manuscript in preparation.

Hopkins, R. O., Gale, S. D., Johnson, S. C., Anderson, C. V., Bigler, E. D., Blatter, D. D., & Weaver, L. K. (1995a). Severe anoxia with and without concomitant brain atrophy and neuropsychological impairments. *Journal of the International Neuropsychological Society*, *1*, 501–509.

Hopkins, R. O., Kesner, R. P., & Goldstein, M. (1995b). Item and order recognition memory in subjects with hypoxic brain injury. *Brain and Cognition*, *27*, 180–201.

Kesner, R. P., & Hopkins, R. O. (2001). Short-term memory for duration and distance in humans: Role of the hippocampus. *Neuropsychology*, *15*, 58–68.

Knowlton, B., Mangels, J., & Squire, L. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.

Knowlton, B., Squire, L., & Gluck, M. (1994). Probabilistic classification learning in amnesia. *Learning and Memory*, *1*, 106–120.

Leng, N., & Parkin, A. (1989). Aetiological variation in the amnesic syndrome: Comparisons using the Brown-Peterson task. *Cortex*, *25*, 251–259.

Myers, C., DeLuca, J., Schultheis, M., Schnirman, G., Ermita, B., Diamond, B., Warren, S., & Gluck, M. (2001). Impaired eyeblink classical conditioning in individuals with anterograde amnesia resulting from anterior communicating artery aneurysm rupture. *Behavioral Neuroscience*, *115*, 560–570.

Parkinsons, R. B., Hopkins, R. O., Cleavinger, B. S., Waver, L. K., Victoroff, J., Foley, J. F., & Bigler, E. D. (2002). White matter hyperintensities and neuropsychological outcome following carbon monoxide poisoning. *Neurology*, *58*, 1525–1532.

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

Press, G., Amaral, D., & Squire, L. (1989). Hippocampal abnormalities in amnesic patients revealed by high-resolution magnetic resonance imaging. *Nature*, *341*, 54–57.

Robbins, T. (1996). Refining the taxonomy of memory. *Science*, *273*, 1353–1354.

Shohamy, D., Onlaor, S., & Gluck, M. (2000). Category learning in Parkinson's disease: The effect of multiple training sessions on probabilistic classification with highly discriminable categories [abstract]. In *Cognitive neuroscience society annual meeting abstracts 2000* (p. 29).

Shohamy, D., Onlaor, S., Myers, C., & Gluck, M. (2001). Category learning in Parkinson's disease: The effect of stimulus presentation on performance [abstract]. In *Cognitive neuroscience society annual meeting abstracts 2001* (p. 36).

Shoqeirat, M., Mayes, A., MacDonald, C., Meudell, P., & Pickering, A. (1990). Performance on tests sensitive to frontal lobe lesions by patients with organic amnesia: Leng & Parkin revisited. *British Journal of Clinical Psychology*, *29*, 401–408.

Spreen, O., & Strauss, E., 1998. *A compendium of Neuropsychological Tests: Administration, norms and commentary*. New York: Oxford University Press.

Squire, L. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In E. Tulving (Ed.), *Memory systems* (pp. 201–231). Cambridge, MA: MIT Press.

Sutherland, R., & Rudy, J. (1989). Configural association theory: The role of the hippocampal formation in learning, memory and amnesia. *Psychobiological*, *17*, 129–144.

Van der Linden, M., Bruyer, R., Roland, J., & Schils, J. (1993). Proactive interference in patients with amnesia resulting from anterior communicating artery aneurysm. *Journal of Clinical and Experimental Neuropsychology*, *15*, 525–536.

Zola-Morgan, S., Squire, L., & Amaral, D. (1986). Human amnesia and the medial temporal region: Enduring memory impairments following a bilateral lesion limited to field CA1 of the hippocampus. *Journal of Neuroscience*, *6*, 2950–2967.