Behavioral/Systems/Cognitive

Declarative Memory, Awareness, and Transitive Inference

Christine Smith[1] and Larry R. Squire[2,3]

[1]Institute for Neural Computation and [2]Departments of Psychiatry, Neurosciences, and Psychology, University of California, San Diego, La Jolla, California 92093, and [3]Veterans Affairs Medical Center, San Diego, California 92161

A characteristic usually attributed to declarative memory is that what is learned is accessible to awareness. Recently, the relationship between awareness and declarative (hippocampus-dependent) memory has been questioned on the basis of findings from transitive inference tasks. In transitive inference, participants are first trained on overlapping pairs of items (e.g., A+B−, B+C−, C+D−, and D+E−, where + and − indicate correct and incorrect choices). Later, participants who choose B over D when presented with the novel pair BD are said to demonstrate transitive inference. The ability to exhibit transitive inference is thought to depend on the fact that participants have represented the stimulus elements hierarchically (i.e., A>B>C>D>E). We found that performance on five-item and six-item transitive inference tasks was closely related to awareness of the hierarchical relationship among the elements of the training pairs. Participants who were aware of the hierarchy performed near 100% correct on all tests of transitivity, but participants who were unaware of the hierarchy performed poorly (e.g., on transitive pair BD in the five-item problem; on transitive pairs BD, BE, and CE in the six-item problem). When the five-item task was administered to memory-impaired patients with damage thought to be limited to the hippocampal region, the patients were impaired at learning the training pairs. All patients were unaware of the hierarchy and, like unaware controls, performed poorly on the BD pair. The findings indicate that awareness is critical for robust performance on tests of transitive inference and support the view that awareness of what is learned is a fundamental characteristic of declarative memory.

Key words: awareness; transitive inference; amnesia; memory; declarative; hippocampus; medial temporal lobe

Introduction

Memory is composed of different abilities that depend on different brain systems (Squire, 1992; Schacter and Tulving, 1994; Eichenbaum and Cohen, 2001). The fundamental distinction is between declarative memory, which depends on the hippocampus and related structures, and a collection of other (nondeclarative) memory abilities that support skill and habit learning, the phenomenon of priming, and other forms of experience-dependent behavior that are expressed through performance rather than recollection.

A characteristic usually attributed to declarative memory is that the acquired knowledge is available to awareness (Eichenbaum, 1997; Gabrieli, 1998). However, the relationship between awareness and declarative memory has recently been questioned on the basis of findings from tasks of transitive inference. In transitive inference, overlapping pairs of items (premise pairs) are first trained (e.g., A+B−, B+C−, C+D−, and D+E−, where + and − indicate correct and incorrect choices). During training, stimulus elements B and D are correct and incorrect equally often. Accordingly, when the novel pair BD is presented at test, reward history itself provides little basis for choosing one

stimulus over the other. However, humans as well as experimental animals sometimes choose B, and the inferential choice of B over D is taken as evidence of transitive inference.

Given that transitive inference performance is dependent on the hippocampal region in experimental animals (Dusek and Eichenbaum, 1997; Buckmaster et al., 2004), and given that conscious awareness for what is learned usually accompanies declarative (hippocampus-dependent) learning in humans, it has seemed reasonable to suppose that awareness of the hierarchical relationship among the stimulus elements (i.e., A>B>C>D>E) should be related to successful performance on transitive inference tasks. Yet, in a recent study of healthy volunteers, who learned a five-item hierarchy, no correlation was found between transitive performance and awareness (Greene et al., 2001). This finding questions the relationship between declarative memory and awareness. Still another study found that above-chance transitive inference performance can occur in unaware individuals (although aware individuals performed much better) (Frank et al., 2005). This finding suggests that the tendency to select one element over another (i.e., a choice of B over D for the BD test pair) need not depend entirely on having aware knowledge about the hierarchical relationship among the elements.

We have reexamined the role of awareness in transitive inference tasks. We assessed performance in participants who were either aware or unaware that the elements in the premise pairs could be arranged in a hierarchy. Experiment 1 involved a five-item transitive inference task, and experiment 2 involved a six-item transitive inference task. In experiment 3, we tested memory-impaired patients with damage thought to be limited to the hippocampal region (CA fields, dentate gyrus, and subicu-

lum) to determine how impaired declarative memory affects transitive inference performance.

Materials and Methods

Experiment 1

Participants. The participants (10 men and 9 women) were undergraduates at the University of California, San Diego who received class credit for participating. They averaged 19.6 years of age (range, 18–29).

Materials and procedure. The stimuli were five characters from the Japanese Hiragana script, as described by Greene et al. (2001). These characters formed an ordered hierarchy such that ゑ＞ま＞か＞め＞せ＞ (or A ＞ B ＞ C ＞ D ＞ E), where ">" describes the relationship "should be selected over." One-half of the participants learned a different ordered hierarchy (か＞せ＞ゑ＞ま＞め＞). Characters were presented on a 15-inch monitor in 48-point font.

The procedure was based on the one described by Greene et al. (2001). Before training, participants were instructed that two figures would appear side by side for 3 s on the screen and they were to select the correct figure. They were also told that at first they would choose by trial and error but that with practice, they might be able to learn which figure was correct. On each trial, participants saw one of four pairs of characters (premise pairs), namely AB, BC, CD, or DE. The correct choice in each premise pair was A+B−, B+C−, C+D−, and D+E−, where + and − indicate the correct and incorrect choices. Participants pressed one button on a keyboard to select the figure on the left and pressed another button to select the figure on the right. Feedback was given after each choice (the word "correct" or "incorrect" appeared on the screen).

Training premise pairs. The four premise pairs were presented in pseudorandom order with the constraint that each premise pair was presented twice within each eight-trial block and no pair could be presented twice in a row. Training continued until participants achieved a sequence of 80 trials in which each of the four premise pairs was correct on at least 18 of 20 trials (90% correct). Premise pairs were presented for 3 s or until the participant made a keyboard press, whichever occurred first. The intertrial interval was 1 s. If participants failed to respond within 3 s, the next trial commenced, and the trial was counted as incorrect. The left–right position of each character in the pair was counterbalanced.

Testing without feedback. Subsequently, the participants were instructed that they should continue to select the same correct figures as they had during training but that feedback would not be given. In the first phase of testing, the four premise pairs (AB, BC, CD, DE) were presented eight times each in pseudorandom order. In the second phase, which was presented immediately after the first, the four premise pairs were each presented four more times but were now intermixed with eight presentations each of two novel (probe) pairs BD (the transitive pair) and AE (the end-anchor pair). Premise pairs and probe pairs were presented for 3 s or until a choice was made, whichever occurred first. If participants failed to respond within 3 s, the trial was counted as incorrect, and the next trial commenced. The left–right position of each character in the pair was counterbalanced.

Assessment of awareness. After testing, participants were asked a series of questions (Table 1) to assess their awareness that (1) a hierarchy existed among the five characters and (2) that the choice of B over D for the transitive probe pair BD could be inferred from knowledge of the hierarchy. To begin, participants were asked to circle the correct character in the pairs BD and AE (Table 1, questions 1 and 2) and to explain why they chose one character over the other. Next, participants were asked to circle the correct character in the novel pairs AC, BE, AD, and CE (Table 1, questions 3–6) and to explain why they chose one character over the other. Finally, participants saw an arrangement of the five characters (Table 1, question 7) and tried to order the characters according to their understanding of how they were related. They were also asked to explain their choice and to describe the strategy they used to arrange the characters. Finally, after completing all of the questions, participants were asked when they came to the strategy that they used to answer the questions: during training, during testing, or while they responded to the questions.

Designation of participants as aware or unaware was based on their responses to the questions in Table 1 and the explanations given for each response. Participants were designated as aware if they indicated in

Table 1. Questions about the hierarchy ゑ ＞ま ＞か ＞め ＞せ (A ＞ B ＞ C ＞ D ＞ E)

1. The following pair of shapes was presented together several times, but you were never told which shape was correct. Circle the shape you believe is correct (guess if necessary).

   め　ま

2. The following pair of shapes was presented together several times, but you were never told which shape was correct. Circle the shape you believe is correct (guess if necessary).

   ゑ　せ

Participants were then asked to specify why they chose one shape over the other in questions 1 and 2.

3. The following pair of shapes was never presented together. Circle the shape you believe is correct (guess if necessary).

   ゑ　か

4. The following pair of shapes was never presented together. Circle the shape you believe is correct (guess if necessary).

   せ　ま

5. The following pair of shapes was never presented together. Circle the shape you believe is correct (guess if necessary).

   め　ゑ

6. The following pair of shapes was never presented together. Circle the shape you believe is correct (guess if necessary).

   か　せ

Participants were then asked to specify why they chose one shape over the other in questions 3–6.

7. Based on your understanding of the relationships among these shapes, arrange the shapes appropriately on the blank lines. Use the numbers provided to stand for the shapes (guess if necessary).

   ₅め
   ₃か
   ₁ゑ
   ₂せ　₄ま

   ____　____　____　____　____
   (1–5)　(1–5)　(1–5)　(1–5)　(1–5)

Participants were then asked to specify why they put the shapes in the order specified above.

The pairs indicated in questions 1–6 are BD, AE, AC, BE, AD, and CE, respectively. The left–right position of characters in questions 1–6 and the arrangement of characters in question 7 were counterbalanced.

their responses and explanations that the characters could be arranged in a hierarchy and that they developed their understanding of the hierarchical relationship among the characters during training and/or during testing. Participants were designated as unaware if their understanding of the hierarchical relationship among the characters developed only as they completed the questions in Table 1 or if they failed to notice the hierarchical relationship among the characters.

The explanations given for each response were particularly important in making designations of awareness and unawareness. Thus, participants could have chosen B over D correctly (Table 1, question 1) because they liked the appearance of B more than D. Such participants would have been designated unaware if they could not then explain the logic of why B should be chosen over D. Conversely, participants would have been designated as aware if they appreciated that there was a hierarchical relationship among the elements, even if the order of the five elements (Table 1, question 7) was not fully correct (e.g., in one case two adjacent characters were misplaced: A ＞ B ＞ D ＞ C ＞ E). Each participant responded in a consistent way across the questions, and there was no ambiguity in designating who was aware and who was unaware.

Participants were also assigned a numerical score from 0 (completely unaware) to 4 (completely aware). The score was assigned depending on how well participants understood that the elements in the premise pairs could be arranged in a hierarchy and how well they understood that the hierarchy could be used to infer logical choices when given the transitive probe pairs. Aware participants were assigned scores of 4 (aware of the correct hierarchy and able to apply the hierarchy to the BD pair) or 3 (aware of the majority of the hierarchy and able to apply the hierarchy to the BD pair). Unaware participants were assigned a score from 0 to 2.

Participants were assigned a score of 2 if they ordered the elements in a mostly correct hierarchy but did not recognize it as a hierarchy and did not recognize how it could be applied to the BD pair. Participants were also assigned a score of 2 if they recognized that the elements might form a hierarchy but ordered the elements mostly incorrectly. Participants were assigned a score of 1 if they first appreciated that the elements could form a hierarchy only while answering the questions in Table 1. Those who never recognized that the elements might form a hierarchy were assigned a score of 0.

Measures derived from training and testing. The measure of principal interest was the accuracy score for each premise pair and each probe pair across the test trials that were given without feedback. To evaluate possible differences in performance during training that might have influenced performance during the test trials, we also calculated three other measures: (1) trials to criterion—this measure was the number of trials needed to reach the performance criterion during the training trials; (2) accuracy scores for each premise pair—this measure was the percentage correct score calculated separately for each premise pair during the course of training; and (3) reward/penalty ratio—during training, the end-anchor elements A and E were always correct or always incorrect, respectively, but the middle elements of the hierarchy (B, C, and D) were correct and incorrect equally often. One might therefore suppose that participants would not be able to rely on differences in the reward histories of choice B and choice D to guide performance when given the transitive probe pair BD. However, this assumption is not valid because choice B and choice D did have different reward histories, relating to the fact that the premise pairs were of different difficulty and were learned at different rates.

Specifically, premise pairs containing end items (AB and DE) were learned faster than premise pairs not containing end items (BC and CD). As a result, choices of B during training were penalized less often than choices of D. The AB pair was learned quickly, so that choices of B had little opportunity to be penalized. Also, B was always reinforced in the slowly learned BC pair. In contrast, choosing D was rewarded and penalized about equally often. The DE pair was learned quickly, and D continued to be rewarded for the remainder of training. Yet D was also penalized during training when it appeared in the slowly learned CD pair. This difference between the reward and penalty histories of B and D could have been used by participants to guide performance when they were given the transitive probe pair BD. Accordingly, we calculated the reward/penalty ratio for the elements B and D by dividing the number of times during training that participants correctly chose B (or D) by the number of times they incorrectly chose B (or D).

Experiment 2

Within a five-item hierarchy, only one transitive probe pair (BD) can be created that does not contain end-anchor elements. When a six-item hierarchy is used, three transitive probe pairs (BD, CE, and BE) can be created, and one therefore need not rely on only one measure to gauge transitive performance.

Participants. The participants (11 men and 8 women) were undergraduates at the University of California, San Diego who received class credit for participating. They averaged 19.2 years of age (range, 18–20).

Materials and procedure. The stimuli were six characters from the Japanese Hiragana script, as described by Frank et al. (2005). These characters formed an ordered six-item hierarchy such that み > そ > の > う > や > ま (or A > B > C > D > E > F), where > describes the relationship "should be selected over." One-half of the participants learned a different ordered hierarchy (う > み > ま > の > そ > や).

The procedure was the same as in experiment 1, except that the characters formed a six-item hierarchy. The correct choice in each premise pair was A+B−, B+C−, C+D−, D+E−, and E+F−, where + and − indicate the correct and incorrect choices.

Training premise pairs. The five premise pairs were presented in pseu-

Table 2. Characteristics of amnesic patients

Patient	Age (years)	Education (years)	WAIS-III IQ	WMS-R				
				Attention	Verbal	Visual	General	Delay
A.B.	67	20	107	87	62	72	54	<50
K.E.	63	13.5	108	114	64	84	72	55
L.J.	67	12	101	105	83	60	69	<50
G.W.	45	12	108	105	67	86	70	<50

WAIS-III IQ, Wechsler Adult Intelligence Scale III; WMS-R, Wechsler Memory Scale-Revised.

dorandom order, as in experiment 1. Each premise pair was presented twice within each 10-trial block, and no pair was presented twice in a row. Training continued until participants achieved a sequence of 100 trials in which each of the five premise pairs was correct on at least 18 of 20 trials (90% correct).

Testing without feedback. Testing was the same as in experiment 1, with the exception that additional premise pairs and probe pairs were presented. In the first phase of testing, the five premise pairs (AB, BC, CD, DE, and EF) were presented eight times each in pseudorandom order. In the second phase, the five premise pairs were each presented four more times, but now intermixed with eight presentations each of four novel (probe) pairs BD, CE, and BE (the transitive pairs) and AF (the end-anchor pair).

Assessment of awareness. The questions asked to assess awareness were similar to those used in experiment 1 but included questions about three transitive probe pairs instead of only one and questions about six novel pairs instead of four. Thus, the first four questions asked about transitive probe pairs BD, CE, and BE, as well as the end-anchor pair AF. The next six questions asked about novel pairs BF, AD, CF, AE, DF, and AC. Finally, participants saw an arrangement of the six characters and tried to order the characters according to their understanding of how they were related. Participants were designated as aware or unaware of the hierarchical relationship among the characters based on their responses to these questions and the explanations given for each response.

Five participants exhibited partial knowledge of the hierarchy and were designated as aware or unaware according to the extent of their knowledge. Two of these participants with partial knowledge were designated aware. One recognized the majority of the hierarchy (A > B > C > D), and one participant recognized the hierarchy and how it could be used to infer the logical choices in transitive probe pairs but did not use the hierarchical logic to answer the other questions. Three other participants with partial knowledge were designated unaware. These three participants misplaced one-half or more of the elements in the hierarchy (i.e., C > A > B > D > E > F, A > C > D > B > E > F, and C > D > A > B > E > F). When the data for these participants were excluded, the findings were the same (see Results).

Experiment 3

Participants. Four memory-impaired patients (three men and one woman) with damage thought to be limited to the hippocampal region (CA fields, dentate gyrus, and subicular complex) participated (Table 2). All of the patients had moderately severe memory impairment. Their scores for copy and delayed (12 min) reproduction of the Rey-Osterrieth figure (Osterrieth, 1944) (maximum score, 36) were 29.3 and 2.5, respectively (controls, 30.3 and 20.6). Paired-associate learning was also severely impaired (10 word pairs/trial for three trials; patients: 0.3, 0.3, and 0.5 word pairs correct/trial; controls: 6.0, 7.6, and 8.9) (Squire and Shimamura, 1986).

Patient A.B. became amnesic in 1976 after an anoxic episode associated with cardiac arrest. K.E. became amnesic in 2004 after an episode of ischemia associated with kidney failure and toxic shock syndrome. L.J. became amnesic in 1988 during a 6-month period with no known precipitating event. Her memory impairment has remained stable since that time. G.W. became amnesic in 2001 after a drug overdose and associated respiratory failure.

For three of the four patients, estimates of medial temporal lobe damage were based on quantitative analysis of magnetic resonance images compared with data for 19 controls (for K.E. and G.W.) or 11 controls (for L.J.) (Gold and Squire, 2005) (Fig. 1). The volumes of the full an-
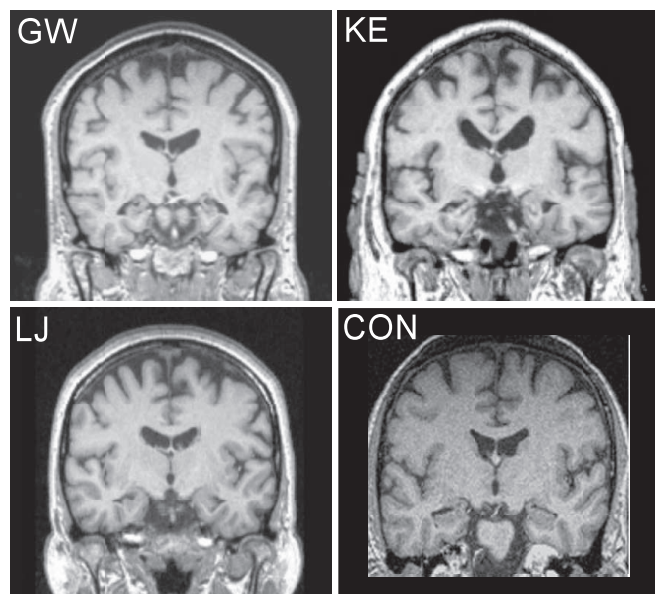
Figure 1. Magnetic resonance images for three of the four memory-impaired patients (G.W., K.E., and L.J.) with damage limited primarily to the hippocampal region and one healthy control (CON). The images are T1-weighted coronal sections at the level of the anterior hippocampus. For all images, the left side of the brain is on the right side of the image. Black triangles on the image for the control, 56 years of age, indicate the hippocampal region. See Materials and Methods for detailed descriptions of the lesions.

Figure 2. Experiment 1. The mean accuracy during testing on a five-item, transitive inference task by healthy volunteers is shown. Premise pair and probe pair performance is shown for participants (mean age, 19.6 years) who were designated as aware or unaware of the hierarchical relationship among the premise pairs. Aware and unaware participants obtained similar accuracy scores on all premise pairs and on the end-anchor probe pair AE. Aware participants performed better than the unaware participants on the transitive probe pair BD (*$p < 0.01$). Error bars indicate SEM. Dashed line, Chance performance.

teroposterior length of the hippocampus and the parahippocampal gyrus were measured using criteria based on histological analysis of healthy brains (Amaral and Insausti, 1990; Insausti et al., 1998). For each patient, the hippocampal and parahippocampal gyrus volumes were divided by the intracranial volume to correct for brain size. K.E., L.J., and G.W. have an average bilateral reduction in hippocampal volume of 49, 46, and 48%, respectively (all values >4 SDs below the control mean). In comparison, the volume of the parahippocampal gyrus was reduced by 17, −8, and 12%, respectively (all values within 2 SDs of the control mean). A.B. is unable to participate in magnetic resonance imaging studies but is thought to have hippocampal damage on the basis of etiology (anoxia) and a neurological examination indicating well circumscribed amnesia. In addition, high-resolution computed tomography images obtained in 2001 were consistent with damage restricted to the hippocampal region (Schmolck et al., 2002).

Additional measurements, based on four controls for each patient, were performed for the frontal lobes, lateral temporal lobes, parietal lobes, occipital lobes, insular cortex, and fusiform gyrus (Bayley et al., 2005a). Volumes of the insular cortex, fusiform gyrus, and each of the major lobes are within ±11% of controls (except for the lateral temporal lobe of patient K.E., which is reduced in volume by 16%). The volumes of each of these anatomical regions are within 1.3 SDs of the control mean.

Thirteen volunteers (nine men and four women) served as controls for the behavioral study. They averaged 57.2 ± 3.7 years of age (patients, 60.5 ± 5.3 years) and had 14.8 ± 0.7 years of education (patients, 14.1 ± 1.7 years).

Materials and procedure. The same Japanese Hiragana characters from experiment 1 were used. One-half of the participants learned one of the ordered hierarchies from experiment 1 (か > せ > ゑ > ま > め or A > B > C > D > E) and one-half learned a different ordered hierarchy (ま > か > せ > め > ゑ). The instructions and procedures were the same as in experiment 1, except that no time limit was imposed for responding and feedback was provided for 2 s instead of 1 s.

Training premise pairs. Before training, participants practiced the task with a pair of geometric shapes (a circle and a triangle; six trials) to become familiar with the feedback screens and with making responses on the keyboard. To make training easier, training proceeded in five progressive phases. For each phase, traini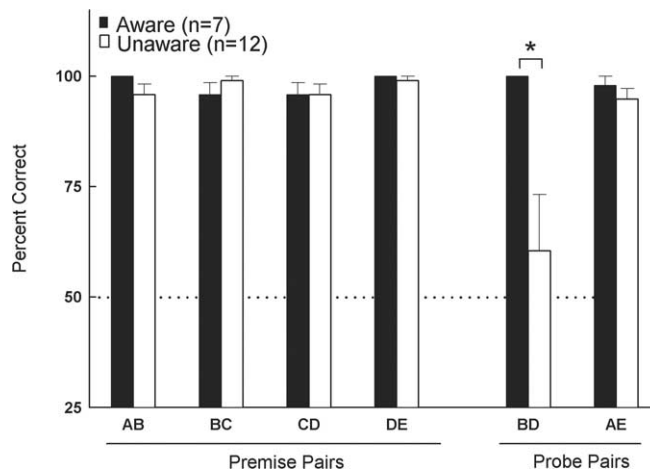ng continued until participants achieved a sequence of 40 trials in which each of the four premise pairs was correct on at least 9 of 10 trials (90% correct). In phase 1, each premise pair (AB, BC, CD, DE) was presented five times consecutively (i.e., 5 × AB, 5 × BC, 5 × CD, 5 × DE, 5 × AB, etc.). After the learning criterion was achieved in phase 1, phase 2 proceeded without interruption. Now, premise pairs were presented twice in a row (i.e., 2 × AB, 2 × BC, 2 × CD, 2 × DE, 2 × AB, etc.) until the learning criterion was reached. In phase 3, premise pairs were presented one time each in sequence (i.e., 1 × AB, 1 × BC, 1 × CD, 1 × DE, 1 × AB, etc.) until the learning criterion was reached. In phase 4, neighboring premise pairs were presented in a series, (i.e., CD, DE, AB, BC, DE, AB, etc.) until the learning criterion was reached. The same premise pair was never presented twice in a row (e.g., AB, BC, BC, CD), and the four neighboring pairs never appeared in sequence (e.g., AB, BC, CD, DE). Finally, in phase 5, the four premise pairs were presented in pseudorandom order with the constraint that each premise pair was presented twice within each eight-trial block and no pair could be presented twice in a row (as in experiment 1).

The controls completed training within a single session. The patients were trained across 2 consecutive days, always beginning with phase 1. On each day, they were permitted to discontinue after 500 trials (or at each 100-trial interval beyond 500 trials). On the second training day, the patients were given no more trials than they had completed on the first training day.

Testing without feedback. None of the patients succeeded at training beyond phase 3, although they did learn some of the premise pairs. Accordingly, at the end of training, all participants were tested exactly as in experiment 1, by giving additional trials without feedback.

Assessment of awareness. Awareness was assessed as in experiment 1.

Results
Experiment 1
Figure 2 shows accuracy scores for premise pairs and probe pairs during testing on the five-item, transitive inference task for participants who were designated as aware ($n = 7$) or unaware ($n = 12$). Aware participants were those who understood that the characters in the premise pairs could be arranged in a hierarchy. Aware and unaware participants obtained similar scores for the premise pairs and for the end-anchor probe pair AE, but scores for the transitive probe pair BD were strikingly different between the two groups ($p < 0.01$). Specifically, aware participants per-

formed nearly perfectly on all of the pairs (all scores above chance; $p < 0.001$), including the transitive probe pair BD. In contrast, the unaware group performed nearly perfectly on the premise pairs and on the end-anchor probe pair AE (all scores above chance; $p < 0.001$) but performed at chance on the transitive probe pair BD ($60.4 \pm 12.8\%$; chance, 50%; $p > 0.40$). Awareness scores were correlated with BD performance ($r = 0.49$; $p < 0.05$) but were not correlated with AE performance ($r = 0.24$; $p > 0.20$).

Although aware and unaware participants performed differently at test, they performed similarly during the course of training. First, aware and unaware participants required a similar number of trials to reach criterion (169.7 ± 22.7 and 186.9 ± 21.6 trials, respectively; $p > 0.60$). Second, when the accuracy scores for each premise pair were averaged across all training trials, the scores of aware and unaware participants were similar across the premise pairs (all $p > 0.10$). Specifically, for pairs AB, BC, CD, and DE, aware participants obtained accuracy scores during training of 84, 74, 71, and 84% correct, respectively; unaware participants obtained scores of 86, 80, 76, and 91% correct, respectively. Finally, for both aware and unaware participants, the reward/penalty ratio during training was higher for B than for D (all $p < 0.05$). Aware participants obtained reward/penalty ratios of 2.24 and 0.68 for B and D, respectively, and unaware participants obtained reward/penalty ratios of 3.76 and 0.74 for B and D, respectively. Because this difference in the reward/penalty ratios was similar for aware and unaware participants (ANOVA, aware/unaware × B/D; $F_{(1,16)} = 1.6$; $p > 0.20$), differences in how often B and D choices were rewarded or penalized cannot account for the marked difference between the two groups in their BD accuracy scores.

In summary, participants who were aware that the elements of the premise pairs could be arranged in a hierarchy performed well when given the transitive probe pair BD. Unaware participants performed poorly. Aware participants were no different from unaware participants on any other measures during training or testing.

Experiment 2

Figure 3 shows accuracy scores for premise pairs and probe pairs during testing on the six-item, transitive inference task for participants who were designated as aware ($n = 6$) or unaware ($n = 12$) at the end of testing. Aware participants were those who understood that the characters in the premise pairs could be arranged in a hierarchy. Aware and unaware participants obtained similar scores for the premise pairs and for the end-anchor probe pair AF, but scores for the transitive probe pairs were different ($p < 0.06$ for BD; $p < 0.05$ for CE and BE). Specifically, aware participants performed nearly perfectly on all of the premise pairs and on all of the transitive pairs except BD [75.0 ± 17.1, 95.8 ± 2.6, and $93.8 \pm 6.3\%$ for BD, CE, and BE respectively; all above chance ($p < 0.001$), except BD ($p = 0.20$)]. Unaware participants performed nearly perfectly on the premise pairs and on the end-anchor probe pair AF (all scores above chance; $p < 0.001$), but they performed poorly on all three transitive probe pairs. Indeed, they performed at chance (50% correct) on the transitive probe pairs BD ($37.5 \pm 9.9\%$; $p > 0.20$) and CE ($62.5 \pm 12.1\%$; $p > 0.30$), and they scored $68.8 \pm 7.8\%$ correct (albeit above chance; $p < 0.05$) on the transitive pair BE. Awareness scores were correlated with overall transitive probe pair performance (awareness vs mean of BD, CE, and BE scores; $r = 0.57$; $p < 0.05$). The separate correlations were $r = 0.47$ for BD ($p < 0.05$), $r = 0.46$ for BE ($p = 0.054$), and $r = 0.44$ for CE ($p = 0.065$).
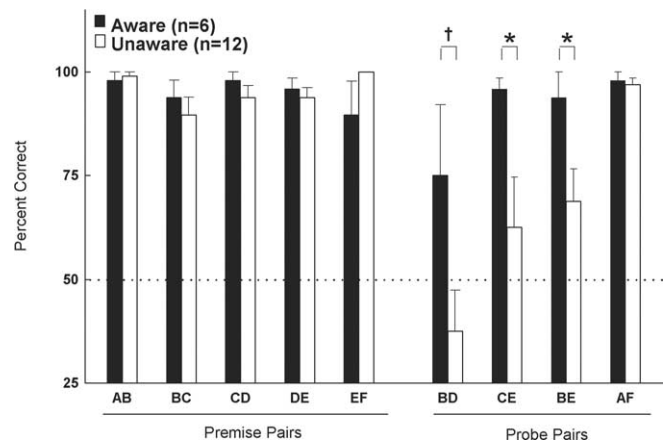

Figure 3. Experiment 2. The mean accuracy during testing on a six-item, transitive inference task by healthy volunteers is shown. Premise pair and probe pair performance is shown for participants (mean age, 19.2 years) who were designated as aware or unaware of the hierarchical relationship among the premise pairs. Aware and unaware participants obtained similar accuracy scores on all premise pairs and on the end-anchor probe pair AE. Aware participants performed better than the unaware participants on the transitive probe pairs BD, CE, and BE ($*p < 0.05$; $^{†}p < 0.06$). Error bars indicate SEM. Dashed line, Chance performance.

Awareness scores were not correlated with AF performance ($r = 0.14$; $p > 0.50$).

Aware and unaware participants performed similarly during the course of training. First, aware and unaware participants required a similar number of trials to reach criterion (279.5 ± 31.5 and 263.0 ± 25.8 trials, respectively; $p > 0.70$). Second, when the accuracy scores for each premise pair were averaged across all training trials, the scores of aware and unaware participants were similar across the premise pairs (all $p > 0.40$). Specifically, for pairs AB, BC, CD, DE, and EF, aware participants obtained accuracy scores during training of 83, 76, 76, 86, and 83% correct, respectively; unaware participants obtained scores of 86, 78, 77, 81, and 89% correct, respectively. Finally, during training, the reward/penalty ratios for the middle elements of the hierarchy were similar between aware and unaware participants. Aware participants obtained reward/penalty ratios for B, C, D, and E of 2.67, 0.62, 0.68, and 0.73, respectively, and unaware participants obtained corresponding reward/penalty ratios for B, C, D, and E of 4.26, 0.64, 0.66, and 0.75. These differences in reward/penalty ratios between the elements that composed the transitive probe pairs BD, BE, and CE could potentially be used by participants to guide performance during testing. However, because these differences were similar for aware and unaware participants [ANOVAs, aware/unaware × B/D (or B/E or C/E); $F_{(1,16)} < 1.2$; $p > 0.20$], differences in how often choices of these elements were rewarded or penalized during training cannot account for differences in how the two groups performed during testing when they were presented with the transitive probe pairs BD, BE, and CE.

Five participants exhibited partial knowledge of the hierarchy and were designated as aware or unaware according to the extent of their knowledge (two had been designated as aware, and three had been designated as unaware). When the data for these participants were excluded, the difference between aware ($n = 4$) and unaware ($n = 9$) participants on the transitive probe trials became more pronounced. Thus, the four aware participants scored 100, 97, and 100% correct on the transitive probe pairs BD, CE, and BE (all scores above chance; $p < 0.001$). The nine unaware participants scored 43, 61, and 68% correct, respectively (all aware vs unaware differences; $p < 0.05$). Transitive probe pair
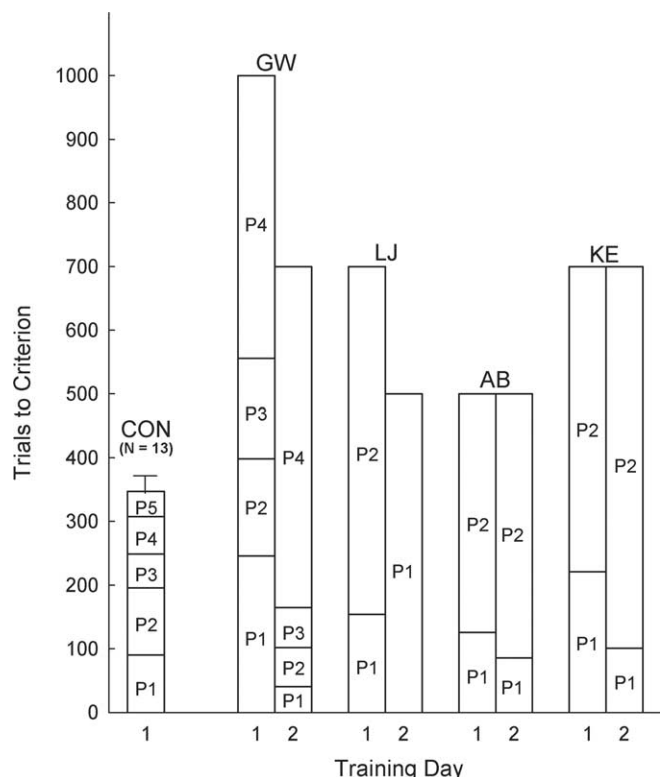
Figure 4. Experiment 3. Trials to criterion during training on a five-item, transitive inference task for healthy controls (CON; mean age, 58.3 years) and memory-impaired patients (GW, LJ, AB, and KE; mean age, 59.7 years). Healthy controls reached criterion (mean ± SEM) within 1 training day (range, 200 – 683 trials). Memory-impaired patients failed to reach criterion during 2 training days. Training progressed in phases (P1–P5). In phase 5, the four premise pairs were presented in pseudorandom order. Participants were required to reach criterion (≥90% accuracy on all premise pairs) within each phase before advancing to the next phase. Error bar indicates SEM.

Figure 5. Experiment 3. The mean accuracy during testing on a five-item, transitive inference task. Premise and probe pair performance is shown for memory-impaired patients (mean age, 60.5 years) and for healthy controls (mean age, 58.3 years) who were designated as aware or unaware of the hierarchical relationship among the premise pairs. Aware and unaware controls obtained similar accuracy scores on all premise pairs and on the end-anchor probe pair AE. Aware controls performed better than unaware controls on the transitive probe pair BD (*$p <$ 0.01). The patients performed well on the premise pair AB and on the end-anchor probe pair AE, but they performed at chance on the other pairs. None of the patients was aware of the hierarchy. Error bars indicate SEM. Dashed line, Chance performance.

scores were at chance levels for these nine participants ($p >$ 0.09). These same aware ($n = 4$) and unaware ($n = 9$) participants performed similarly during the course of training. The two groups required a similar number of trials to reach criterion and obtained similar accuracy scores for the premise pairs during training (all $p > 0.30$). The reward/penalty ratios were also similar for the two groups [ANOVAs, aware/unaware × B/D (or B/E or C/E); $F_{(1,11)} < 0.40; p > 0.50$].

In summary, participants who were aware that the elements of the premise pairs could be arranged in a hierarchy performed well on the transitive probe pairs, and participants who were unaware performed poorly. Aware and unaware participants performed similarly on all other measures during training and testing.

Experiment 3

Controls ($n = 13$) learned the premise pairs in a single session in 347.5 ± 43.9 trials (range, 200–683 trials). In contrast, memory-impaired patients ($n = 4$) were unable to learn the premise pairs to criterion, even after two sessions of training on consecutive days (Fig. 4). Nevertheless, except for pair BC, patients did have some success at learning. Thus, during the second day of training, scores averaged 94, 61, 80, and 89% correct, respectively, for pairs AB, BC, CD, and DE. At the same time, these values overstate how well the patients performed, because the scores mostly reflect performance during the easier stages of training (phases 1 and 2) (Fig. 4). In comparison, controls averaged 94, 90, 91, and 97% correct across all five phases of training that included earlier, easy
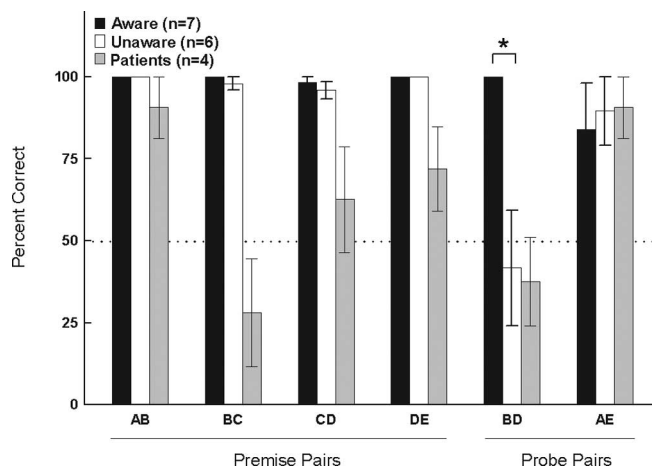
phases and later, more difficult phases. Finally, both the patients and the controls exhibited the end-anchor effect in that the two premise pairs containing end items (AB and DE) were learned faster than the inner premise pairs (BC and CD). In summary, although patients were impaired at learning the premise pairs, they did learn a little, and the relative difficulty of the premise pairs was similar for patients and controls (i.e., the end-anchor pairs were easier than the inner pairs).

Both the patients and the controls were administered test trials at the end of training. Figure 5 shows accuracy scores for premise pairs and probe pairs during testing for patients, as well as for controls who were designated as aware ($n = 7$) or unaware ($n = 6$). Aware individuals were those who understood that the characters in the premise pairs could be arranged in a hierarchy. Aware and unaware controls obtained similar scores for the premise pairs and for the end-anchor probe pair AE, but scores for the transitive probe pair BD were markedly different between aware and unaware groups ($p < 0.05$). Specifically, aware participants performed nearly perfectly on all of the pairs, including the transitive probe pair BD [all scores above chance ($p < 0.001$), except AE ($p = 0.053$)]. In contrast, the unaware group performed nearly perfectly on the premise pairs and on the end-anchor probe pair AE (all scores above chance; $p < 0.05$) but performed at chance on the transitive probe pair BD ($41.7 \pm 17.6\%$; chance, 50%; $p > 0.60$). Awareness scores of the controls were correlated with BD performance ($r = 0.62; p < 0.05$) but not with AE performance ($r = -0.05; p > 0.80$).

Like the participants in experiment 1, aware and unaware controls also performed similarly during the course of training. First, aware and unaware controls required a similar number of trials to reach criterion in each training phase ($p > 0.40$) and required a similar number of total trials (340.3 ± 59.6 and 356.0 ± 71.0 trials, respectively; $p > 0.80$). Second, when the accuracy scores for each premise pair were averaged across all training trials, the scores of aware and unaware controls were similar across premise pairs (all $p > 0.08$). Specifically, for pairs AB, BC, CD, and DE, aware controls obtained accuracy scores during training of 94, 91,

88, and 96% correct, respectively; unaware controls obtained scores of 93, 88, 93, and 98% correct, respectively. Finally, for both aware and unaware controls, the reward/penalty ratio during training was higher for B than for D. Aware controls obtained reward/penalty ratios of 9.84 and 0.85 for B and D, respectively, and unaware controls obtained reward/penalty ratios of 9.20 and 0.91 for B and D, respectively. Because the difference in the reward/penalty ratios was similar for aware and unaware controls (ANOVA, aware/unaware × B/D; $F_{(1,10)} = 0.03$; $p > 0.80$), differences in how often B and D choices were rewarded or penalized cannot account for the difference between the two groups in their BD accuracy scores.

As might be expected for a group that did not complete training, the memory-impaired patients performed poorly on most of the test pairs. Specifically, performance on pairs BC, CD, and DE did not exceed chance levels (all $p > 0.10$). Nevertheless, performance was good and similar to control performance on the premise pair AB and on the end-anchor probe pair AE (scores were above chance; $p < 0.05$). Performance on the transitive probe pair BD was at chance ($p > 0.40$), although this score is not very meaningful because the patients did not successfully learn the training pair BC. Finally, none of the patients was aware that the characters in the premise pairs could be arranged in a hierarchy.

In summary, memory-impaired patients had difficulty learning the premise pairs and did not reach criterion, even after 2 training days. The controls performed like the younger participants in experiment 1. Participants who were aware that the elements of the premise pairs could be arranged in a hierarchy performed well when given the transitive probe pair BD. The unaware participants performed poorly. Aware and unaware participants performed similarly on all other measures during training and testing.

Discussion

In three experiments, we assessed transitive inference performance as a function of awareness that the elements of the task could be arranged in a hierarchy. In experiment 1, which involved a five-item transitive inference task, aware participants performed well on the transitive probe pair BD, but unaware participants performed at chance. Experiment 2 used a six-item transitive inference task (with three transitive probe pairs) and yielded similar results. In experiment 3, memory-impaired patients were given a five-item transitive inference task. On the transitive probe pair BD, controls performed like the participants in experiments 1 and 2. Specifically, aware controls performed well, and unaware controls performed poorly. Memory-impaired patients had difficulty learning the premise pairs and did not reach criterion, even after 2 training days.

Our results are consistent with previous findings that aware participants performed better than unaware participants on measures of transitive inference (Martin and Alsop, 2004; Frank et al., 2005). Note, however, that differences between aware and unaware transitive inference performance in these previous studies could have reflected more than differences in awareness. For example, in one study (Martin and Alsop, 2004), aware and unaware participants did differ in their transitive performance at test, but they also differed in their premise pair performance at test, suggesting that the two groups had not learned the premise pairs equally well. The present results demonstrate that, even when participants who were aware of the hierarchy and those who were unaware learned at the same rate and performed similarly on a number of other measures, aware participants still

performed much better than unaware participants on tests of transitive inference.

In marked contrast to our results, Greene et al. (2001) found no correlation between awareness and transitive inference performance. In their study, an awareness score (1–5) was assigned depending on when during the questioning, and to what extent, awareness of the hierarchy could be expressed. In their experiment 1, 17 of 22 participants had moderate to high levels of awareness. Transitive inference scores were also high (87% correct across all 22 participants). Given that both awareness scores and transitive inference performance were high, there may have been too little variability for a correlational analysis to yield a meaningful value [for another example, see Titone et al. (2004)]. It is also difficult to interpret the findings from two additional experiments in which the transitive pair BD was presented repeatedly during training, before the premise pairs themselves had been fully learned (Greene et al., 2001). Transitive inference performance is most directly assessed in tests given after the premise pairs have been learned.

In our experiment 2, the pattern of performance exhibited by unaware participants on the three transitive probe pairs (BD, CE, and BE) deserves comment. The finding was that performance on the transitive pair BE, which was closest to the end-anchors A and F, was above chance. Performance was intermediate for transitive pair CE, and performance was lowest for transitive pair BD. The linear trend across these three transitive pairs was significant ($F_{(1,11)} = 11.3$; $p < 0.01$). This graded performance has been reported previously in both humans and rats (Van Elzakker et al., 2003; Frank et al., 2005) and has been taken to suggest that reinforcement-based learning strategies can be involved in transitive inference. According to this idea [value transfer theory (von Fersen et al., 1991)], an associative strength gradient is established during training across the elements in the premise pairs. As outlined in this previous work, the gradient is created because some of the associative strength accrued for the end-anchor elements is shared with the immediately neighboring elements, and this associative strength becomes weaker with increasing distance from the end anchors. Our finding in experiment 2 of differential performance across the transitive pairs by unaware participants is consistent with the idea that associative strengths (and presumably nondeclarative memory) can account for this aspect of transitive inference performance.

At the same time, unaware participants did not exhibit measurable transitive inference performance in our experiments 1 and 3. Although the value transfer theory ordinarily predicts above-chance performance in these cases, Frank et al. (2005) pointed out that participants can also adopt explicit, albeit incorrect, strategies or rules that compete with the benefit accrued from differences in associative strengths. Consistent with this idea, unaware participants in our three experiments exhibited considerable variability in BD performance and frequently described incorrect reasons for their choices. The important point is that the contribution of reinforcement-based (unaware) learning to performance was small and that aware participants consistently outperformed unaware participants by a considerable margin.

It is also important to note that some participants designated as unaware nevertheless exhibited good transitive inference performance. For example, in our experiment 1, although the mean transitive inference score for the unaware group was 60.4% (not different from chance), the scores nevertheless ranged from 0 to 100% correct. High scores can occur because some participants, although unaware of the hierarchical relationship among the

stimulus elements, nevertheless choose B over D because, for example, they like it more or because they have some incorrect impression about the structure of the task. Accordingly, it will always be the case that unaware participants will be represented among those participants who perform well.

This phenomenon was described previously (Siemann and Delius, 1993, 1996; Delius and Siemann, 1998) and taken in support of the idea that participants who are unaware can exhibit transitive inference and that awareness provides no benefit to performance. Yet, transitive inference scores were compared only for those aware and unaware participants who exhibited good transitive inference performance. In this high-performing subgroup, aware participants performed no better than unaware participants. However, the important question concerns not only the number of aware and unaware participants who perform well on the transitive pairs but also the number who perform poorly. When we performed this calculation for the one study in which sufficient data were presented (Siemann and Delius, 1993) (eight aware and seven unaware participants performed well; one aware and eight unaware participants performed poorly), the effect of awareness was significant (Fisher's exact test; $p < 0.05$). These considerations support the findings of our study and others (Martin and Alsop, 2004; Frank et al., 2005) that consistent and correct performance on measures of transitive inference requires awareness of the hierarchical relationship among the stimulus elements.

In experiment 3, memory-impaired patients performed poorly on the test of transitive inference (probe pair BD), presumably as a result of their difficulty in learning the premise pairs. These results differ from findings for monkeys with lesions of the entorhinal cortex and rats with lesions of the fornix or perirhinal/entorhinal cortex, who learned the premise pairs in a similar number of trials as controls and then exhibited a selective impairment on the transitive probe pair BD (Dusek and Eichenbaum, 1997; Buckmaster et al., 2004).

The difference between the performance of memory-impaired patients and experimental animals is likely attributable to differences in how the premise pairs are learned. Humans appear to approach this task declaratively, and they attempt to memorize the correct element in each premise pair. Nondeclarative memory might ultimately support this kind of learning, but only if much more extended training were given (Bayley et al., 2005b). In any case, if premise pairs cannot be learned in the time available, there is no possibility of learning about the hierarchy and no possibility of exhibiting transitive inference.

In contrast, experimental animals appear to learn by engaging nondeclarative (habit) memory (Mishkin and Petri, 1984). For example, monkeys with lesions of the hippocampal region learned both the pattern discrimination and the eight-pair concurrent discrimination tasks at a normal rate (Buffalo et al., 1998; Teng et al., 2000) but were impaired at learning these tasks after lesions of the caudate nucleus (Divac et al., 1967; Teng et al., 2000; Fernandez-Ruiz et al., 2001). Accordingly, we suggest that experimental animals with hippocampal lesions or lesions of related structures succeed at learning premise pairs because the pairs are acquired by habit learning and the learning depends on the caudate nucleus. The same animals subsequently fail on the transitive inference test pair (Dusek and Eichenbaum, 1997; Buckmaster et al., 2004), because they cannot perform the hippocampus-dependent computations that are needed specifically for transitive inference performance

[see accounts by Rapp et al. (1996), Dusek and Eichenbaum (1997), and Frank et al. (2003)].

In three experiments, we found that performance on transitive inference tasks benefits markedly from awareness that the elements among the premise pairs can be arranged in a hierarchy. This finding extends to both older and younger participants and to five-item and six-item transitive inference tasks. Nondeclarative, reinforcement-based accounts of transitive inference may explain how unaware participants can sometimes exhibit above-chance performance on some transitive pairs (see experiment 2). Nevertheless, it is clear that awareness of the hierarchical relationship among the stimulus elements is related to successful task performance and that awareness is critical for robust performance on tasks of transitive inference. These findings support the view that awareness of what is learned is a fundamental characteristic of declarative (hippocampus-dependent) memory.

References

Amaral DG, Insausti R (1990) Hippocampal formation. In: The human nervous system (Paxinos G, ed), pp 711–755. San Diego: Academic.

Bayley PJ, Gold JJ, Hopkins RO, Squire LR (2005a) The neuroanatomy of remote memory. Neuron 46:799–810.

Bayley PJ, Frascino JC, Squire LR (2005b) Robust habit learning in the absence of awareness and independent of the medial temporal lobe. Nature 436:550–553.

Buckmaster CA, Eichenbaum H, Amaral DG, Suzuki WA, Rapp PR (2004) Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. J Neurosci 24:9811–9825.

Buffalo EA, Stefanacci L, Squire LR, Zola SM (1998) A reexamination of the concurrent discrimination learning task: the importance of anterior inferotemporal cortex, area TE. Behav Neurosci 112:3–14.

Delius JD, Siemann M (1998) Transitive responding in animals and humans: exaptation rather than adaptation? Behav Processes 42:107–137.

Divac I, Rosvold HE, Szwarcbart MK (1967) Behavioral effects of selective ablation of the caudate nucleus. J Comp Physiol Psychol 63:184–190.

Dusek JA, Eichenbaum HB (1997) The hippocampus and memory for orderly stimulus relations. Proc Natl Acad Sci USA 94:7109–7114.

Eichenbaum H (1997) Declarative memory: insights from cognitive neurobiology. Annu Rev Psychol 48:547–572.

Eichenbaum H, Cohen NJ (2001) From conditioning to conscious recollection: memory systems of the brain. New York: Oxford UP.

Fernandez-Ruiz J, Wang J, Aigner TG, Mishkin M (2001) Visual habit formation in monkeys with neurotoxic lesions of the ventrocaudal neostriatum. Proc Natl Acad Sci USA 98:4196–4201.

Frank MJ, Rudy JW, O'Reilly RC (2003) Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. Hippocampus 13:341–354.

Frank MJ, Rudy JW, Levy WB, O'Reilly RC (2005) When logic fails: implicit transitive inference in humans. Mem Cognit 33:742–750.

Gabrieli JDE (1998) Cognitive neuroscience of human memory. Annu Rev Psychol 49:87–115.

Gold JJ, Squire LR (2005) Quantifying medial temporal lobe damage in memory-impaired patients. Hippocampus 15:79–85.

Greene AJ, Spellman BA, Dusek JA, Eichenbaum HB, Levy WB (2001) Relational learning with and without awareness: transitive inference using nonverbal stimuli in humans. Mem Cognit 29:893–902.

Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkanen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. Am J Neuroradiol 19:659–671.

Martin N, Alsop B (2004) Transitive inference and awareness in humans. Behav Processes 67:157–165.

Mishkin M, Petri HL (1984) Memories and habits: some implications for the analysis of learning and retention. In: Neuropsychology of memory (Squire LR, Butters N, eds), pp 287–296. New York: Guilford.

Osterrieth PA (1944) Le test de copie d'une figure complexe [The test of copying a complex figure]. Archives de Psychologie 30:206–356.

Rapp PR, Kansky MT, Eichenbaum H (1996) Learning and memory for

hierarchical relationships in the monkey: effects of aging. Behav Neurosci 110:887–897.

Schacter D, Tulving E (1994) Memory systems 1994. Cambridge, MA: MIT.

Schmolck H, Kensinger E, Corkin S, Squire LR (2002) Semantic knowledge in patient HM and other patients with bilateral medial and lateral temporal lobe lesions. Hippocampus 12:520–533.

Siemann M, Delius JD (1993) Implicit deductive responding in humans. Naturwissenschaften 80:364–366.

Siemann M, Delius JD (1996) Influences of task concreteness upon transitive responding in humans. Psychol Res 59:81–93.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 99:195–231.

Squire LR, Shimamura AP (1986) Characterizing amnesic patients for neurobehavioral study. Behav Neurosci 100:866–877.

Teng E, Stefanacci L, Squire LR, Zola SM (2000) Contrasting effects on discrimination learning after hippocampal lesions and conjoint hippocampal-caudate lesions in monkeys. J Neurosci 20:3853–3863.

Titone D, Ditman T, Holzman PS, Eichenbaum H, Levy DL (2004) Transitive inference in schizophrenia: impairments in relational memory organization. Schizophr Res 68:235–247.

Van Elzakker M, O'Reilly RC, Rudy JW (2003) Transitivity, flexibility, conjunctive representations, and the hippocampus. I. An empirical analysis. Hippocampus 13:334–340.

von Ferson L, Wynne CDL, Delius JD, Staddon JER (1991) Transitive inference in pigeons. J Exp Psychol Anim Behav Process 17:334–341.