

Learning strategies in amnesia

Maarten Speekenbrink, Shelley Channon and David R. Shanks

Department of Psychology
University College London

Abstract

Previous research suggests that early performance of amnesic individuals in a probabilistic category learning task is relatively unimpaired. When combined with impaired declarative knowledge, this is taken as evidence for the existence of separate implicit and explicit memory systems. The present study contains a more fine-grained analysis of learning than earlier studies. Using a dynamic lens model approach with plausible learning models, we found that the learning process is indeed indistinguishable between an amnesic and control group. However, in contrast to earlier findings, we found that explicit knowledge of the task structure is also good in both the amnesic and the control group. This is inconsistent with a crucial prediction from the multiple-systems account. The results can be explained from a single system account and previously found differences in later categorization performance can be accounted for by a difference in learning rate.

Over the past decades, neuropsychology has shown increasing interest in probabilistic category learning. In a widely-used categorization task, known as the “weather prediction” task (Knowlton, Squire, & Gluck, 1994), the objective is to predict the weather (sunny or rainy) on the basis of four cues (tarot cards with different geometric patterns). Introducing the task as a variant of the medical diagnosis task (Gluck & Bower, 1988), Knowlton et al. (1994) found that, relative to controls, early categorization performance was not impaired in amnesia. Amnesic patients with medial temporal lobe or diencephalic lesions showed apparent normal learning, despite their severe declarative memory problems. One explanation for this finding, favoured by many authors (e.g. Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005; Gabrieli, 1998; Knowlton et al., 1994; Knowlton, Mangels, & Squire, 1996; Poldrack & Rodriguez, 2004), is in terms of multiple memory systems. Knowlton et al. (1994) argued that early learning in multiple cue tasks is mainly dependent on procedural (implicit) memory, which is not impaired in amnesia. This hypothesis gained further support from the “double dissociation” between categorization performance and declarative memory. Knowlton et al. (1996) found that, while amnesic patients showed relatively unimpaired early categorization performance, they had impaired declarative (explicit) memory for the task episode. Parkinson’s patients, who are assumed to have impaired procedural but unimpaired declarative memory, showed relatively impaired categorization performance from the outset, but did show accurate memory of the testing episode (Knowl-

ton et al., 1996).

While this evidence is appealing, it shows weaknesses upon closer examination. The analyses only focus on the effect of learning (categorization performance), but neglect the underlying learning process. At present, it is unclear how individuals learn in the weather prediction task. Moreover, the declarative knowledge tapped in the experiments referred to irrelevant task features, rather than to the object of learning. The purpose of the present paper is to submit the learning behaviour and declarative knowledge to a more sensitive analysis. We analysed participants' (amnesic and control) learning on a trial-by-trial basis and investigated how they integrated information to arrive at their predictions. We compared their performance to normative learning models, in order to gain further insight into possible differences between amnesic and normal participants. By applying a dynamic lens model approach, we captured the progression of participants' implicit response strategies. Finally, by tapping participants' explicit knowledge regarding the task structure and their categorization strategies, we investigated whether there is a dissociation between procedural and declarative memory in probabilistic category learning.

Evidence for a double dissociation in probabilistic category learning.

The empirical evidence for a double dissociation between procedural and declarative memory concerns two groups of individuals; one group shows normal categorization performance but impaired declarative memory of the task, while the second group shows impaired categorization performance but unimpaired declarative knowledge of the task. In our short review of the evidence for this dissociation, we limit discussion to research conducted with the weather prediction task (for a more complete overview, the reader is referred to Gabrieli, 1998; Ashby & Maddox, 2005).

It is an established finding that, compared to controls, amnesic individuals show unimpaired early learning (Knowlton et al., 1994, 1996; Eldridge, Masterman, & Knowlton, 2002; Reber, Knowlton, & Squire, 1996; but see Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004, for a different result). More specifically, what has been found is that categorization performance of amnesic groups does not significantly differ from performance of matched controls in the first 50 trials of the weather prediction task. Since the performance of both groups does differ significantly from chance performance, both groups show learning, in an apparently similar manner. However, when asked about certain features of the task, such as where the cards were placed on a computer screen, or the number of elements of the geometric patterns on the cards, amnesic patients show very poor memory of these task features (Knowlton et al., 1996; Eldridge et al., 2002). Hence, there appears to be unimpaired learning, but impaired declarative memory in amnesia. Learning, so the argument goes, must thus be based on a procedural memory system which is independent of the medial temporal lobes. Parkinson's patients, on the other hand, show impaired categorization performance from the outset (Knowlton et al., 1996; Shohamy, Myers, Onlaor, & Gluck, 2004), while their knowledge of task features is normal (Knowlton et al., 1996; Sage et al., 2003). This complements the evidence, in that participants can show accurate declarative memory of task features, while having an impaired ability to learn to categorize the stimuli.

But declarative knowledge of these task features is irrelevant (Lovibond & Shanks, 2002). To show that learning is based on a procedural memory system inaccessible to

conscious recollection, we should consider knowledge of the object of learning (i.e., the contingencies between cards and weather). This task knowledge is to be distinguished from another relevant type of declarative knowledge: self-insight. Self-insight refers to an individual's knowledge of how (s)he uses the cues in order to predict the outcome (Lagnado, Newell, Kahan, & Shanks, 2006). In previous research, these two types of declarative knowledge have often been conflated, and it is not clear whether the claimed dissociation between procedural and declarative memory refers to a dissociation between classification performance and task knowledge, classification performance and self-insight, or classification performance, task knowledge and self-insight. To our knowledge, the effect of amnesia on self-insight has not been investigated previously, while there has been only one study (Reber et al., 1996) addressing amnesia and task knowledge. Reber et al. (1996) found that amnesic patients lacked explicit knowledge of the probabilities in the task, while controls' judged probabilities were quite close to the objective values. However, in this study, participants were only shown 50 trials. Although this may be enough to ensure better-than-chance performance, it is hardly enough to gain adequate knowledge of the contingencies between cards and weather. Some of the cue patterns for which participants were asked to rate the probability of the outcome were shown only once; at least in a statistical sense, estimating a probability from a single observation is highly problematic. In this light, it is not remarkable that the amnesic participants lacked knowledge. What is more surprising is that the control subjects knew as much as they did.

This brings us to a related problem: the results mainly concern early performance. After more trials, categorization performance has been found to differ between amnesic and control participants (Knowlton et al., 1994). Knowlton et al. explain this later divergence by assuming that, as the task progressed, control participants may have formed strategies which depend on declarative memory. Amnesic patients would not have been able to develop such strategies. Shohamy et al. (2004) found that performance of Parkinson's patients did increase steadily with training, although not to the level of control subjects. They offer a similar explanation for this finding, namely that Parkinson's patients were apparently able to develop declarative learning strategies. However, alternative explanations are also consistent with this pattern. For instance, different patient groups could simply learn at lower rates than controls (e.g., Kinder & Shanks, 2001; McClelland & Rumelhart, 1986), which would result in more marked differences in performance as the task progresses.

Neuroimaging studies have sought more direct evidence for the involvement of different memory systems in the weather prediction task. Poldrack et al. (2001) compared performance of healthy participants in two versions of the task: the standard feedback-based version, and a paired-associate learning version, in which participants were asked to memorize cue-outcome pairings. The latter version was designed to engage declarative memory. In a subsequent test phase, participants in both conditions showed equal categorization performance, but differences in neuronal activity: relative to a baseline task, participants who learned the feedback-based version showed more activity in the caudate nucleus, and less activity in the medial temporal lobe (MTL), while participants who learned the paired associate version showed the opposite pattern. Furthermore, a second experiment with just the feedback-based version showed that the MTL was initially active, and the caudate nucleus inactive, but this pattern was quickly reversed, with the MTL becoming inactive, and the caudate nucleus active. On the basis of these findings, Poldrack et al. (2001) propose

a competition between MTL- and striatum-based memory systems in probabilistic category learning. According to this account, the MTL acquires flexible, relational knowledge, while the striatum acquires inflexible stimulus-response associations. Competition is necessitated by the assumed fundamental incompatibility of these types of information. Foerde, Knowlton, and Poldrack (2006) tried to increase reliance on implicit memory by assigning participants a secondary task. They showed that performance was not different under single or dual task conditions, but that the secondary task impaired declarative knowledge of the cue-outcome contingencies. Moreover, single and dual task learning resulted in different neuronal activity; performance on items learned under single task conditions was correlated with activity in the right hippocampus, while performance on items learned under dual task conditions was correlated with activity in the putamen.

While these neuroimaging studies have shown the involvement of both the MTL and striatum in the weather prediction task, and that their relative involvement can be modulated by task demands, whether they provide evidence for separate implicit and explicit memory/learning systems is debatable. Implicit memory entails a lack of explicit knowledge, and while Foerde et al. (2006) found impaired declarative knowledge under dual task conditions, Newell, Lagnado, and Shanks (in press) failed to replicate this finding in a similar experiment. Moreover, contrary to Foerde et al., the dual task condition performed significantly below the single task condition. In a second experiment, Newell et al. (in press) found that participants in a feedback- and observation-based version of the weather prediction task did not differ in declarative knowledge, which is unexpected if participants relied on different memory systems. As such, it is questionable whether the manipulations were successful in engaging implicit learning, and thus, whether the differences in neuronal activity reflect different memory systems. By itself, that task demands engage distinct regions of the brain does not imply the involvement of different memory systems (Sherry & Schacter, 1987).

To summarize, the evidence for multiple memory systems is weak. Claims of dissociable implicit and explicit memory systems rely on strong assumptions, such as the inaccessibility to conscious recollection. Valid tests of these assumptions have been rare and inconclusive. The empirical data do not rule out the possibility that individuals learn in qualitatively similar ways, but that a neurological disorder (amnesia, or Parkinson's disease) affects the rate of learning. This quantitative difference should affect procedural and declarative memory alike (given that both pertain to the same object, i.e., the contingencies in the environment). In contrast to this single system explanation, a multiple memory systems account would predict that individuals learn in qualitatively different ways (depending on the memory system involved), or that performance on implicit and explicit tests relies on different memory systems. To test these different hypotheses, we need to investigate the learning process directly, and employ valid tests of procedural and declarative memory.

Learning strategies

The strategies employed in solving probabilistic category learning tasks were investigated by Gluck, Shohamy, and Myers (2002). Based on participants' self-reports, they formulated three broad learning strategies and compared how well these described participants' categorization behaviour. Overall, participants seemed to adopt a singleton strategy, in which they predicted the optimal outcome for singleton patterns (patterns in which only

a single cue is present), and guessed for other patterns. It was also apparent that strategy use changed during the task. Towards the end of the task, a large proportion of participants seemed to adopt a multi-cue strategy, in which predictions were based on all available cues. The paradigm has later been used to distinguish learning strategies of different patient groups (e.g. Hopkins et al., 2004; Shohamy et al., 2004). In particular, Hopkins et al. (2004) found that amnesic patients overwhelmingly used a singleton strategy, while controls showed more variety in their strategies.

There is reason to question the validity of the strategies defined by Gluck et al. (2002). A main problem of the multi-cue strategy as formulated by Gluck et al. is that it is a deterministic strategy (i.e., for each cue pattern for which the probability of an outcome is greater than .5, a participant is taken to predict that outcome with probability 1). This is also known as a maximising strategy. It has been shown often that participants fail to follow this (optimal) maximising strategy, even when they know the structure of the task. Participants often show a probability matching strategy, in which the probability of a particular response is close to the probability that this response is correct (Shanks, Tunney, & McCarthy, 2002). Hence, the poor fit of the multi-cue strategy could be the result of probability matching, rather than a failure to integrate the information of all cues. Indeed, as Lagnado et al. (2006) showed, by including a multi-cue/probability matching strategy, the evidence for the singleton strategy vanishes, as most participants are characterised by this multi-match strategy. Meeter, Myers, Shohamy, Hopkins, and Gluck (2006) also found that the preponderance of the singleton strategy could be attributed to it accounting for random and below chance responses better than the other strategies.

For our purposes, a more important problem with the strategies is that they pertain more to response than learning processes. This distinction may seem subtle, but its importance has been stressed before (e.g. Massaro & Friedman, 1990; Friedman & Massaro, 1998; Kitzis, Kelley, Berg, Massaro, & Friedman, 1998). While the learning process involves acquiring knowledge regarding the contingencies in the environment, the response process involves the use of this knowledge to form predictions. Strategies such as the singleton or multi-cue strategy require knowledge of the contingencies between cues and outcome, which can only be acquired by learning through repeated observation of cues and outcome. In other words, the *learning* process is not captured by the models, which only predict the *response* given that the knowledge has been acquired. The learning process may or may not be dependent on the response process (for instance, in multiple cue learning, the feedback governing learning is independent of the response given, while in reinforcement learning, the response determines the subsequent reinforcement which in turn drives learning), but the response process is usually considered to be dependent on the knowledge acquired through learning. For probabilistic category learning, we take the view that the learning process is primary, with the response process being dependent on it (but not vice versa). Although the effects of the learning and response process are confounded in participants' responses, advances have been made to distinguish the two (e.g., Friedman, Massaro, Kitzis, & Cohen, 1995; Friedman & Massaro, 1998; Kitzis et al., 1998).

To conclude, the validity of the strategy-based account is unclear. The strategies are partly based on self-report, but this is inconsistent with the hypothesis that probabilistic category learning is a form of implicit learning. Explicit learning on the other hand should result in a good correspondence between reported and employed strategy. However, Gluck

et al. (2002) found a general mismatch between reported strategy and best-fitting strategy on the basis of formal modelling. Also, the models proposed are all of the form ‘given pattern X the participant always responds A ’, while research has consistently shown that participants are much more variable in their responses. Finally, the strategy-based account does not inform us how participants learn, although an improved model which allows strategy switches (Meeter et al., 2006), is a step in the right direction. Hence, to study learning strategies in amnesia, or anywhere else, we should look for a different approach. Here, we develop formal models of learning in probabilistic categorization tasks, and then apply them to learning data from an amnesic and control group.

Learning process

In multiple cue probability learning (MCPL) the objective is to predict the outcome of a criterion variable y on the basis of J cues $\mathbf{x} = (x_1, \dots, x_J)^\top$, where \top denotes the transpose. For example, in the Weather Prediction Task, the objective is to predict the state of the weather y , which can be sunny ($y = 0$) or rainy ($y = 1$), on the basis of four tarot cards (x_j), each of which can be presented ($x_j = 1$) or not ($x_j = 0$). The cues and outcome are probabilistically related, and predictions should be based on the conditional distribution $P(y|\mathbf{x})$. By repeated observation of paired observations (y, \mathbf{x}) , it is possible to learn the particulars of this conditional distribution, and hence to arrive at optimal predictions. The optimal prediction for a given cue pattern \mathbf{x} is simply to predict $r = 1$ if $P(y = 1|\mathbf{x})$ is larger than $P(y = 0|\mathbf{x})$, and $r = 0$ otherwise.

We will focus on models in which the conditional probability is approximated by a function of the cue values. When y is a binary variable, as it is here, a useful choice is the logistic regression function

$$P(y = 1|\mathbf{x}) \simeq [1 + \exp(-\mathbf{w}^\top \mathbf{x})]^{-1}, \quad (1)$$

in which the regression weights $\mathbf{w} = (w_1, \dots, w_J)^\top$ can be estimated from repeated observations of cues and outcome. The true values of these regression weights¹ reflect how well the outcome can be predicted from the cues, and we will refer to them as *cue validity* weights \mathbf{v} . We define learning as the change in \mathbf{w} (the current inference of \mathbf{v}) over trials t , where the change is in the direction of \mathbf{v} .

We now present two models which function in this way. Both models are models for on-line learning, by which we mean that the change in \mathbf{w} from one trial to the next is determined solely by the information (\mathbf{x}_t and y_t) presented at that trial. This is to be contrasted with models for batch learning, in which the change in weights is determined by the information from previous trials as well (usually *all* previous observations). The two models presented next are normative models in the sense that they can be guaranteed to arrive at \mathbf{v} in the limit as $t \rightarrow \infty$.

Associative learning

The first model is an associative model, in which the weights can be interpreted as associations between cues x_j and outcome y , which increase or decrease in strength based

¹By true values we mean the population values of the regression weights, i.e. the values of \mathbf{w} for which the approximation in Equation 1 is closest to the real value of $P(y|\mathbf{x})$.

on the given information. At each trial t , weights are determined by changing them in the direction which minimizes the ‘cross-entropy’ error (C. M. Bishop, 1995)

$$e_t = - \left\{ y_t \ln f(\mathbf{w}^\top \mathbf{x}_t) + (1 - y_t) \ln [1 - f(\mathbf{w}^\top \mathbf{x}_t)] \right\}, \quad (2)$$

in which $f(\cdot)$ is the logistic function of Equation 1. This leads to the following recursive relation

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta_t [y_t - f(\mathbf{w}_t^\top \mathbf{x}_t)] \mathbf{x}_t, \quad (3)$$

in which η_t is the learning rate parameter which determines the size of the change in the direction that minimises the error. When the learning rate meets certain conditions², it can be shown that \mathbf{w}_t converges to \mathbf{v} as $t \rightarrow \infty$ (Robbins & Monro, 1951). A simple scheme obeying these conditions is $\eta_t = \eta/t$. Another option, which does not guarantee convergence, is a constant learning rate $\eta_t = \eta$, for which Equation 3 becomes the LMS rule (Gluck & Bower, 1988).

It should be noted that this associative learning model is identical to a single layer feedforward neural network with a logistic activation function. As such, it appears identical to the model used by Gluck and Bower (1988). However, while the models are indeed very similar, there is a crucial distinction in that, due to the logistic activation and error function employed, this model learns the probabilities in the environment directly. Gluck and Bower’s model used a linear activation function, incorporating a logistic transformation only ‘after the fact’ to model probabilities of response. Moreover, we will allow for decreasing learning rate, while the LMS learning rule used by Gluck and Bower only uses a constant learning rate.

Bayesian learning

The second model we consider instantiates Bayesian learning. Bayesian models have become increasingly popular, both in statistics and psychology (e.g. Anderson, 1991; Chater, Tenenbaum, & Yuille, 2006). In the Bayesian framework, learning concerns the distribution of parameters, rather than point estimates of the parameter values. This is done through the recursive relation

$$g_t(\mathbf{w} | \mathbf{x}_{1:t}, y_{1:t}) = f(y_t | \mathbf{x}_t, \mathbf{w}) g_{t-1}(\mathbf{w} | \mathbf{x}_{1:t-1}, y_{1:t-1}) / K, \quad (4)$$

in which the likelihood function $f(\cdot)$ is the logistic function in Equation 1, g_t the posterior density of \mathbf{w} , g_{t-1} the prior density of \mathbf{w} , and the constant K (which is computed by integrating the right hand side of Equation 4 with respect to \mathbf{w}) ensures that g_t is a proper probability density (i.e., it integrates to 1). Starting from a prior density $g_0(\mathbf{w})$, learning consists of determining g_t at each trial t from \mathbf{x}_t , y_t and g_{t-1} .

While the associative model accounts for decreasing weight of new observations by a decreasing learning rate, the Bayesian model does so implicitly. Due to the nature of Bayesian learning, the posterior becomes more and more peaked around a value of \mathbf{w} when more observations come in. Because of this, the relative effect of a new observation on the posterior distribution decreases. This implication is somewhat analogous to the computation of a mean: the relative weight of each observation when computing the mean from 10

²These conditions are (1) $\lim_{t \rightarrow \infty} \eta_t = 0$, (2) $\sum_{t=1}^{\infty} \eta_t = \infty$, and (3) $\sum_{t=1}^{\infty} \eta_t^2 < \infty$.

observations is 1/10, while the relative weight is 1/100 when computing the mean from 100 observations.

Response process

When the outcome (and hence the response) is categorical, it is an established observation that $P(r|\mathbf{x}_t)$ is usually not identical to $P(y|\mathbf{x}_t)$, even though the participants may have learned the latter probability correctly (Shanks et al., 2002). In fact, probability matching, where the probabilities are approximately equal, is not the optimal strategy for predicting the outcome. Maximising the probability of a correct prediction requires $P(r = j|\mathbf{x}_t) = 1$ when $P(y = j|\mathbf{x}_t) > .5$. $P(r|\mathbf{x}_t)$ usually lies between probability matching and this maximising strategy. In any case, it is clear that individuals differ in the way in which they act upon learned contingencies. A particularly useful way to capture such individual differences is to assume

$$P(r = 1|\mathbf{x}_t) = \frac{P(y = 1|\mathbf{x}_t)^\lambda}{P(y = 1|\mathbf{x}_t)^\lambda + P(y = 0|\mathbf{x}_t)^\lambda}, \quad (5)$$

in which $P(y = 1|\mathbf{x}_t)$ represents the model predicted probability of Equation 1 and $\lambda > 0$ is a response scaling parameter (e.g. Friedman & Massaro, 1998; Nosofsky & Zaki, 1998; Zaki, Nosofsky, Stanton, & Cohen, 2003). When $\lambda = 1$, the response process is probability matching, while values of $\lambda > 1$ indicate deviations towards maximising. If λ lies between 0 and 1, this indicates ‘undershooting’ so that the individual under-utilizes the learned contingencies. When the relation in Equation 5 holds, it can be shown that $\mathbf{u} = \lambda\mathbf{w}$ (see the Appendix), where \mathbf{u} contains the *cue utilization* weights, the regression weights for the logistic regression model for the responses (Equation 1 when the left hand side is replaced with $P(r = 1|\mathbf{x})$). In other words, if a participant learns as the model prescribes, then the utilization weights are a linear function of the inferred cue validity weights (the weights of the model fitted to the outcome). The slope of this linear function is equal to λ .

Lens Model approach

The learning models and response process describe how participants respond if they exactly follow the learning process. However, participants’ actual behaviour may deviate from these normative models. For instance, participants may not focus on all the cues to the same extent. As such, the utilization weights will deviate from the inferred cue validity weights of an ideal observer. In order to investigate such discrepancies, we relied on a lens model analysis. The lens model approach has proven a valuable tool to capture participants’ implicit judgement policies in multiple cue tasks (for a review, see Cooksey, 1996; Goldstein, 2004). The general idea is to fit two regression models, one regressing the outcome y on the cues \mathbf{x} , and the other regressing the response r on the cues. With this approach, one can analyse performance (usually defined as the correlation between y and r) as a function of the particulars of the environment (cues and outcome) and judgement system (cues and response). In particular, the regression weights of the environment (\mathbf{v} , the true, and \mathbf{w} , the inferred cue validity weights) are compared to those from the model of the judgement system (\mathbf{u} , the cue utilization weights). Recently, the approach has been applied to study dynamical learning processes by applying the lens model to a moving

window of trials (Kelley & Friedman, 2002; Lagnado et al., 2006). This “rolling regression” technique provides a sequence of estimates of the utilization weights and cue validities. By comparing \mathbf{u}_t , the cue utilization weights, to \mathbf{w}_t , the inferred cue validity weights of an ideal observer, one can investigate how the participants learned relative to how well they could have learned.

The approach adopted here is similar in the sense that the same model is fitted to both outcome and response. However, rather than using a rolling regression analysis, we use the on-line learning models described previously. In this way, we try to capture more realistic learning processes, while maintaining the ability to compare participants’ performance to that of an ideal learning mechanism, provided with the same information.

Method

Participants

Nine participants (6m, 3f) with memory disorders took part in the study (see Table 1). Four of these were diagnosed with alcoholic Korsakoff’s syndrome. Four had developed memory problems from suspected temporal lobe damage following encephalitis, and MRI radiological investigation in two of these cases confirmed medial temporal lobe involvement. One had a right-sided medial temporal lobe tumour, confirmed by radiological investigation. Sixteen control participants also took part in the study (7m, 9f). To be included in the study, participants had to be between 18 and 75 years of age, fluent in English, within the normal range on the Test of Reception of Grammar (D. V. M. Bishop, 1989), and have an IQ score of 85 or above. Exclusion criteria included significant physical or psychiatric illness (with the exception of the primary diagnosis for the amnesic group), hydrocephalus and dementing conditions, and expressive or receptive aphasia. The two groups of participants did not differ significantly ($p > 0.1$) in age, years of education, or Wechsler Abbreviated Scale of Intelligence (WASI) IQ (Wechsler, 1999).

All participants were assessed on the Wechsler Memory Scale III (Wechsler, 1997), in addition to the experimental task (Table 1). The amnesic group performed significantly more poorly than the control group on each of these measures ($p < .0001$ except for the Auditory Recognition Delayed Index and Working Memory Index, where $p < .01$), and degree of memory impairment in individual amnesic participants ranged from mild to marked. Participants were also tested on the Tower, Trail-Making, and Colour Word Interference subtests of the Delis-Kaplan Executive Function System (DKEFS) test (Delis, Kaplan, & Kramer, 2001). Results are given in Table 2. The amnesic group did not differ significantly from the control group ($p > .05$) on the Tower and Trail-Making subtests, or on three of the four components of the Colour-Word Interference subtest; they did differ significantly on the Inhibition/switching component of the latter subtest.

Materials

The Weather Prediction task was identical to that used by Lagnado et al. (2006). The stimuli presented to participants were taken from a set of four cards, each with a different geometric pattern (squares, diamonds, circles, triangles). The task consisted of a total of 200 trials, on each of which participants were presented with a pattern of one, two or three cards. Each trial was associated with one of two outcomes (Rain or Fine), and overall

Table 1: Etiology, sex, age, years of education, IQ, and Wechsler Memory Scale III scores for the amnesic and control group. WMS scales: Auditory Immediate (AI), Auditory Delayed (AD), Auditory Recognition Delayed (ARD), Auditory Recognition (VI), Visual Immediate (VI), Visual Delayed (VD), Immediate Memory (IM), General Memory (GM), and Working Memory (WM).

	etiology*	sex	age	education	IQ	Wechsler Memory							
						auditory			visual				
						AI	AD	ARD	VI	VD	IM	GM	WM
Amnesic													
A1	Enceph.	F	51	16	102	62	61	70	91	78	71	64	88
A2	Enceph.	F	50	13	100	105	102	90	81	81	93	89	79
A3	Kors.	M	51	11	87	89	83	75	88	88	86	79	83
A4	Kors.	M	49	16	125	105	105	115	81	91	93	103	131
A5	Enceph.	M	57	13	118	111	86	110	68	75	89	84	99
A6	MTL t.	M	75	16	106	83	74	75	61	68	67	67	91
A7	Enceph.	M	50	16	113	77	83	90	81	84	74	82	88
A8	Kors.	F	56	16	99	62	58	80	57	65	51	60	81
A9	Kors.	M	52	11	103	59	55	65	61	53	51	47	83
Mean			54.56	14.22	105.88	83.67	78.56	85.56	74.33	85.56	75.00	75.00	91.44
SD			8.14	2.22	11.32	20.24	18.20	17.40	12.72	17.40	17.03	17.03	16.02
Control													
Mean			56.75	13.94	104.50	110.38	113.44	109.06	99.88	109.06	106.69	108.00	109.31
SD			10.95	2.14	14.43	11.60	9.57	15.94	13.73	14.83	11.25	9.87	14.83

*Enceph. = Encephalitis, Kors. = alcoholic Korsakoff's syndrome, MTL t. = MTL tumour.

Table 2: Means and standard deviations of the Delis-Kaplan Executive Function System scales.

	Amnesic		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Tower	10.89	2.47	11.27	2.15
Trail making				
motor speed	9.33	3.16	10.13	1.73
visual scanning	9.00	1.50	10.31	2.60
number sequencing	10.11	1.97	10.50	2.45
letter sequencing	10.00	1.87	11.00	2.36
number-letter switching	10.22	2.77	11.25	1.92
Colour Word Interference				
colour naming	8.44	3.32	10.07	3.08
colour word reading	10.67	2.12	11.20	2.83
colour inhibition	8.44	4.50	11.07	2.05
inhibition/switching	7.56	4.16	11.13	2.56

these two outcomes occurred equally often. The pattern frequencies and probabilities of the outcome were identical to those in Gluck et al. (2002, Experiment 2) and subsequent other studies (e.g., Hopkins et al., 2004; Lagnado et al., 2006), and are shown in Table 1, along with the probability of the outcome for each of these 14 patterns. The learning set was constructed so that each card was associated with the outcome with a different probability. For example, the probability of rain was .2 over all the trials on which the squares card (card 1) was present, .4 for trials on which the diamonds card (card 2) was present, .6 for trials on which the circles card (card 3) was present, and .8 for trials on which the triangles card (card 4) was present. Thus, two cards were predictive of rainy weather, one strongly (card 4), one weakly (card 3), and two cards were predictive of fine weather, one strongly (card 1), one weakly (card 2). Overall, participants experienced identical pattern frequencies (order randomized for each participant), but the actual outcome for each pattern was determined probabilistically (so experienced outcomes could differ slightly across participants). The position of the cards on the screen were held constant within participants, but counterbalanced across participants.

Procedure

The procedure of the experiment was identical to that in Lagnado et al. (2006), and a detailed description can be found there. Participants were given on-screen instructions, after which they moved onto the first block of 50 trials. On each trial a specific pattern of cards (selected from Table 3) was displayed on the screen. Participants were then asked to predict the weather on that trial, by clicking on the corresponding button (RAINY or FINE). Once they had made their prediction, participants received immediate feedback as to the actual weather on that trial, and whether they were correct or incorrect.

At the end of each block of fifty trials participants answered two different sets of test questions. In the probability test participants were asked to give probability ratings for

Table 3: Learning environment.

Pattern	Cards present	Total	P(pattern)	P(fine pattern)
A	0001	19	0.095	0.895
B	0010	9	0.045	0.778
C	0011	26	0.13	0.923
D	0100	9	0.045	0.222
E	0101	12	0.06	0.833
F	0110	6	0.03	0.500
G	0111	19	0.095	0.895
H	1000	19	0.095	0.105
I	1001	6	0.03	0.500
J	1010	12	0.06	0.167
K	1011	9	0.045	0.556
L	1100	26	0.13	0.077
M	1101	9	0.045	0.444
N	1110	19	0.095	0.105
Total		200	1.00	

each of the four cards. For each card they were asked for the probability of rainy vs. fine weather: ‘On the basis of this card what do you think the weather is going to be like?’ They registered their rating using a continuous slider scale ranging from ‘Definitely fine’ to ‘Definitely rainy’, with ‘As likely fine as rainy’ as the midpoint. In the importance test participants were asked how much they had relied on each card in making their predictions: ‘Please indicate how important this card was for making your predictions’. They registered their rating using a continuous slider scale ranging from ‘Not important at all’ to ‘Very important’, with ‘Moderately important’ as the midpoint.

Results

Learning Performance

It has been customary to analyse participants’ performance in terms of the proportion of optimal responses (e.g. Hopkins et al., 2004; Knowlton et al., 1994, 1996; Shohamy et al., 2004). However, such a measure of performance does not distinguish between optimal responses to highly predictive versus weakly predictive patterns³. Instead, we assessed participant’s performance through the score statistic

$$S = \frac{1}{T} \sum_{t=1}^T r_t P(y = 1|\mathbf{x}_t) + (1 - r_t)[1 - P(y = 1|\mathbf{x}_t)], \quad (6)$$

³Consider two patterns \mathbf{x}_1 and \mathbf{x}_2 . The probability of Rain given pattern 1 is .92, and given pattern 2 it is .56. We would consider a non-optimal response (Sun) for pattern 1 as worse than a non-optimal response (Sun) for pattern 2. This difference is not taken into account in the proportion of optimal responses. In other words, only considering optimal responses might not distinguish between someone who always predicts Rain to pattern 1 and Sun to pattern 2, and someone who always predicts Sun to pattern 1 and Rain to pattern 2. However, the expected number of correct predictions is much higher for the first than for the second person.

in which $r_t = \{0, 1\}$ is a participant's response on trial t , and T is the total number of trials. This score is the average of the probability that a response results in a correct prediction, and gives a more reliable estimate of the expected performance than the proportion of correct predictions.

The mean scores for each block of fifty trials are shown in Figure 1. Across the task participants steadily improved in their ability to predict the outcome. A 1 between (group) \times 1 within (block) ANOVA showed a significant linear trend for block, $F(1, 23) = 11.19$, $MSe = 5.94 \times 10^{-3}$, $p < .01$. There was no significant main effect of group, $F(1, 23) = 2.49$, $MSe = 2.68 \times 10^{-2}$, $p = .13$, nor a significant interaction between group and the trend for block, $F(1, 23) = .074$, $MSe = 5.95 \times 10^{-3}$, $p = .79$. Hence, although Figure 1 shows that the mean scores of the amnesic group did lie consistently below those of the controls, there appeared to be no significant difference between the amnesic and control group in learning performance⁴.

While we found no significant differences in performance between the amnesic and control group, overall performance was strongly related to the measures of visual memory from the Wechsler Memory Scale. In particular, we found a significant correlation between performance and immediate visual memory, $r(23) = .56$, $p < .01$, and delayed visual memory, $r(23) = .45$, $p < .05$, while correlations between immediate and delayed auditory memory were not significant, $r(23) = .37$ and $r(23) = .33$ respectively.

Learning models

We fitted several versions of the learning models (Table 4). For the associative models, a first distinction is whether the learning rate decreased over trials (models Decr/x/x in Table 4), or whether it was constant (models Const/x/x). A second distinction is whether the learning rate parameter was fixed, or free. For the fixed parameter case, the parameter values were chosen to be optimal⁵. For the free parameter case, the value of the learning rate parameter was estimated from the data, by maximising the likelihood of responses, using the inferred cue validity weights as utilization weights. To investigate the extent of individual variation in learning rate, we compared models with a single learning rate for all participants (models x/Single/x), a different learning rate for the amnesic and control group (models

⁴ We repeated the analysis with the usual outcome measure (proportion of optimal responses). For the amnesic group, the average percentages of optimal responses were 52.67 ($SD = 9.95$), 59.33 ($SD = 10.1$), 65.11 ($SD = 15.78$) and 64.00 ($SD = 16.7$) for blocks 1-4 respectively. For the control group, the average percentages of optimal responses were 61.25 ($SD = 10.98$), 64.25 ($SD = 16.44$), 70.25 ($SD = 13.91$) and 71.38 ($SD = 12.94$) for blocks 1-4 respectively. The ANOVA had very similar results, with a significant linear trend for block, $F(1, 23) = 11.92$, $MSe = 149.51$, $p < .01$. There was no significant main effect of group, $F(1, 23) = 2.01$, $MSe = 485.22$, $p = .17$, nor an interaction between group and block, $F(1, 23) = 0.02$, $MSe = 149.51$, $p = .88$. The similarity of results is not surprising, as the score statistic of Equation 6 correlates highly with the proportion of optimal responses, $r = .98$. However, as the score statistic is more reliable, and power is an issue, we prefer it to the other measure.

⁵For the model with decreasing learning rate, $\eta_t = \eta/t$, the parameter value was $\eta = 5.66$, which is equal to 1 over the largest eigenvalue of the expected Hessian matrix, evaluated at the optimal parameter vector. This is the optimal value, in the sense that it is the highest learning rate that does not result in divergence (LeCun, Bottou, Orr, & Miller, 1998). For the model with a constant learning rate, $\eta_t = \eta$, the optimal parameter can not be determined analytically. The parameter value $\eta = .5$ was determined heuristically, allowing the weights to reach their optimal values in the space of 200 trials, while not resulting in too much variance around these optimal values.

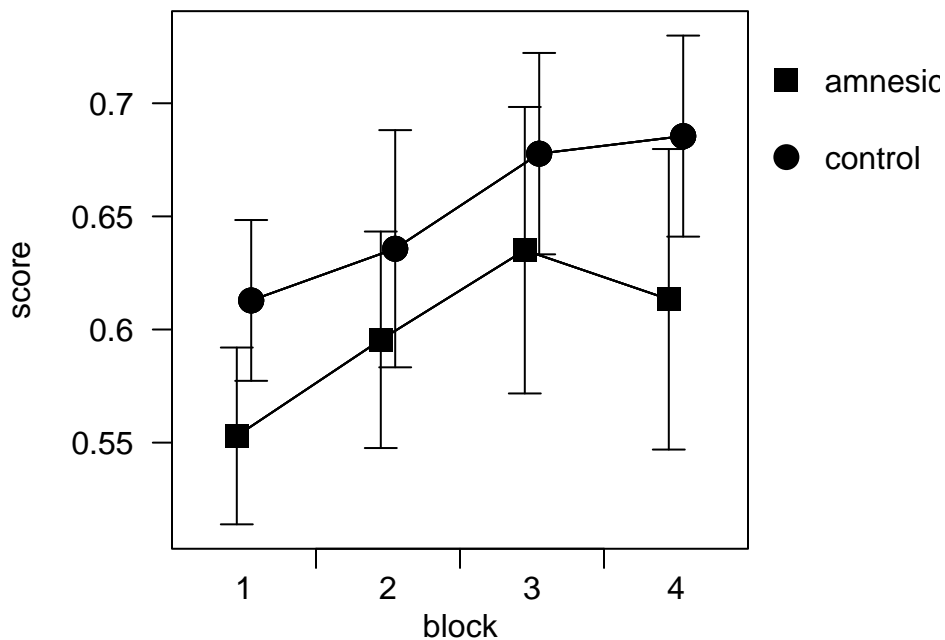


Figure 1. Averages and 95% confidence intervals of the performance scores by block, for the amnesic (circles) and control group (triangles). Curves are slightly displaced on the horizontal axis to enhance visualisation of the confidence intervals.

x/Group/x), and a different learning rate for each participant (models x/Ind/x). When there is little variation in learning rate, the model with a single learning rate parameter should show a similar fit to the data as the others. If there is little variation within the two groups, the model with a group dependent learning rate should show a similar fit to the model with a different learning rate for each individual. A final distinction concerns the response process. For all models, we fitted a version which assumed participants responded by probability matching (models x/x/Match, for which the response scaling parameter was fixed to $\lambda = 1$), and, to allow for deviations from probability matching, a version in which the response scaling λ was estimated as a free parameter for each individual (models x/x/Free). We fitted two versions of the Bayesian learning model. The first version (model Bayes/Match) assumed probability matching ($\lambda = 1$), while the second (model Bayes/Free) allowed for individual variation in the response process.

The weights for the associative model were derived through the recursive relation of Equation 3. The starting weights at $t = 0$ were all set to 0. The joint distribution of the weights in the Bayesian model was approximated by a discrete multidimensional grid with seven equally spaced grid points for each weight (the joint distribution was thus approximated by a discrete probability distribution over $7^4 = 2401$ possible values). The prior distribution g_0 was set to a uniform distribution, and the values of the weights were estimated as posterior means. All other parameters (η 's and λ 's) were estimated by maximum likelihood estimation. To compare the fit of models we used the Bayesian Information Criterion (BIC, Schwarz, 1978).

The results of the model fitting are given in Table 4. The BIC indicates that the associative model with a constant, individually varying, learning rate, and probability matching for the response process (model Const/Ind/Match) is preferred to all other models. Although the same model with individual variation in the response process (model Const/Ind/Free) had a higher likelihood, the BIC shows that this increase in likelihood does not justify the 25 additional parameters it required. The associative model with a constant fixed learning rate, but individual variation in response scaling (model Const/Opt/Free), had an identical number of free parameters to model Const/Ind/Match, but did not fit the data as well. Hence, individual differences in learning rate appear more important than differences in response scaling. Furthermore, pairwise comparisons between the associative models with constant and decreasing learning rate show that the former always fitted the data better than the latter (i.e., the fit of Const/Opt/Match was better than Decr/Opt/Match, the fit of Const/Opt/Free was better than Decr/Opt/Free, etc.). This strongly indicates that participants learned at a constant, rather than decreasing rate. Finally, for the Bayesian model, the response scaling parameter was crucial, since model Bayes/Match fitted worse than a null-model, which assumes that each response is completely random (i.e., made by flipping an unbiased coin).

Interestingly, there appears to be a difference in the learning rate between the amnesic and control group. As expected, amnesic individuals had a lower learning rate than controls. The mean learning rate was .11 ($SD = .16$) in the amnesic group, and .23 ($SD = .20$) in the control group, which is a marginally significant difference, Mann-Whitney $U = 43.5$, $p = .054$ (one-sided). A similar pattern occurs for the Bayesian model with response scaling, where there was a significant difference in the mean of the response scaling parameter for the amnesic ($M = .31$, $SD = .21$) and control group ($M = .54$, $SD = .30$), Mann-Whitney $U = 39$, $p < .05$ (one-sided). Note that these response scaling parameters were rather low (usually, they are larger than 1), indicating ‘undershooting’ (the value of the response scaling parameter is less than that for probability matching).

Lens model analysis

As described in the introduction, the lens model approach consists of fitting two models simultaneously, one to the environment (cues and outcome), and one to the response system (cues and response). The parameters of the two models are then compared. The weights for the normative models should approach the cue validity weights of $v_1 = -2.10$, $v_2 = -.58$, $v_3 = .58$, $v_4 = 2.10$. As for the cue utilization weights, they may be lower than (‘undershooting’), the same as (‘matching’), or higher than (‘overshooting’) the optimal weights.

As the associative learning model with a constant learning rate described participants’ responses best, we will focus on this model in the lens model analysis. Also, a constant learning rate has the added advantage that the model is more sensitive to the changes in response probability resulting from a dynamical learning process. To estimate cue utilization, we fitted the associative model with a constant learning rate fixed at $\eta = .5$ (model Const/Opt/Match), to participants’ responses. Figure 2 shows the trial-by-trial estimates of the cue utilization weights for each participant. In most cases, participants learned the weights for the strong cards rather well. We see that for most participants, the weights for card 1 and 4 diverged quickly, and in the expected direction (negative weights for card 1

Table 4: Fit measures for associative model with decreasing learning rate (Decr/x/x), associative model with constant learning rate (Const/x/x), and Bayesian model (Bayes/x). Learning rate (rate) and response scaling (scaling) were either fixed (value given) or free parameters (greek symbol, with possible subscript to indicate that the parameter varies over groups (grp) or individuals (ind)). $\log(L)$ indicates log likelihood, (#par) the number of free parameters, and $\text{BIC} = -2 \times \log(L) + (\text{\#par}) \log(N)$, where N is the total number of observations. Best fitting model is indicated in bold font. The BIC of a null model, which assigns probability .5 to each observation, is 6931.47.

Model	rate	scaling	$-\log(L)$	#par	BIC
Associative, decreasing learning rate					
Const/Opt/Match	5.66/ t	1	3813.92	0	7627.84
Const/Opt/Free	5.66/ t	λ_{ind}	3035.42	25	6283.78
Const/Single/Match	η/t	1	3165.60	1	6339.72
Const/Single/Free	η/t	λ_{ind}	3022.51	26	6274.76
Const/Group/Match	η_{grp}/t	1	3100.26	2	6339.72
Const/Group/Free	η_{grp}/t	λ_{ind}	2968.54	27	6167.04
Const/Ind/Match	η_{ind}/t	1	3026.48	25	6265.88
Const/Ind/Free	η_{ind}/t	λ_{ind}	3007.04	50	6439.95
Associative, constant learning rate					
Decr/Opt/Match	.5	1	3349.83	0	6699.66
Decr/Opt/Free	.5	λ_{ind}	2875.36	25	5963.66
Decr/Single/Match	η	1	3041.10	1	6090.74
Decr/Single/Free	η	λ_{ind}	2871.04	26	5963.53
Decr/Group/Match	η_{grp}	1	3020.72	2	6058.48
Decr/Group/Free	η_{grp}	λ_{ind}	2869.21	27	5968.38
Decr/Ind/Match	η_{ind}	1	2865.26	25	5943.44
Decr/Ind/Free	η_{ind}	λ_{ind}	2845.45	50	6116.76
Bayesian					
Bayes/Match	—	1	3566.67	0	7133.34
Bayes/Free	—	λ_{ind}	2917.60	25	6048.13

and positive for card 4). From trial 100 onwards, the weights of card 4 were usually close to the optimal value. The same holds for the weight of card 1, although there is evidence that this card was overweighted by some participants (most noticeably C1, C5, C10 and A3). As for the weak cards, the pattern is less clear, although for most participants, the weights diverged in the expected directions (w_3 being higher than w_2). Also, note that some participants (e.g. C1, C6, C9 and A3) overweighted card 3. Participant C10 showed the best overall performance. As can be seen in Figure 2, the cue utilization weights of this participant diverged quickly and show the appropriate ordering. Participant C9 was the worst overall performer and the cue utilization weights reflect this, showing that this participant implicitly took the weak cards (2 and 3) to be indicative of Rain, and the strong cards (1 and 4) as indicative of Sun. Finally, participant A6 deserves special mention. The cue utilization weights indicate that this participant hardly relied on any cue in making predictions. Not surprisingly, this was also one of the worst performing participants.

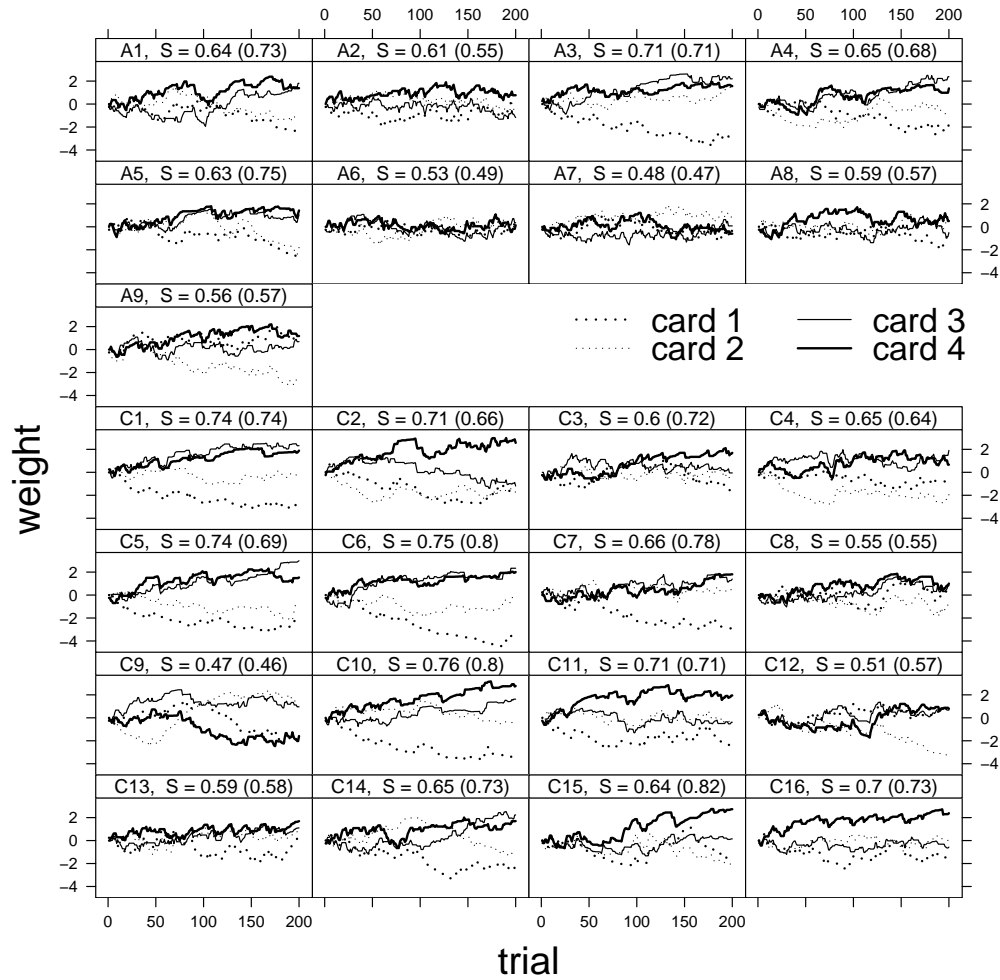


Figure 2. Cue utilization weights for each participant, estimated by model Const/Opt/Match. A1-A9 are amnesic and C1-C16 are control participants. The S besides participant id denotes the overall performance score as in Equation 6. The score in the final block (trials 150-200) is given in parentheses. Strongly predictive cards (1 and 4) have thick lines, weakly predictive cards (2 and 3) thin lines. Cards predictive of Sun (1 and 2) have broken lines, cards predictive of Rain (3 and 4) solid lines.

While the cue utilization profiles are informative, especially when interpreted in relation to the optimal cue weights, they do not provide information as to how well the participants were performing relative to how well they could have performed if they followed a normative learning model precisely. For this, we need to compare the cue utilization weights to the inferred validity weights of an ideal observer, i.e. the weights derived from the normative learning model fitted to the outcome. Figure 3 shows the average utilization and ideal observer weights. As this figure shows, the cue utilization weights are somewhat closer to the ideal observer weights for the controls than for the amnesic participants. This difference between groups appears especially marked for the strongly predictive cards (card 1 and card 4). To test for group differences, we computed a 2 (group) \times 4 (card) \times 2 (source: ideal observer or utilization) ANOVA for the weights, with repeated measures on the last two factors and a linear and quadratic trend for trial. This analysis showed a significant main effect of card, $F(3, 69) = 160.92$, $MSe = 84$, $p < .001$, and significant interactions between card and source, $F(3, 69) = 20.65$, $MSe = 59.6$, $p < .001$, card and linear trend, $F(3, 69) = 44.63$, $MSe = 32.7$, $p < .001$, card and the quadratic trend, $F(3, 69) = 10.40$, $MSe = 12.11$, $p < .001$, as well as a significant three-way interaction between card, source and quadratic trend, $F(3, 23) = 5.10$, $MSe = 8.40$, $p < .01$. This last effect indicates a stronger quadratic trend for the ideal observer than for the utilization weights, which is due to the ideal observer weights converging to the optimal weights in the space of 200 trials, while the utilization weights showed no clear convergence. There were no significant main or interaction effects involving group; interaction between group, card, and source, $F(3, 69) = 1.69$, $MSe = 59.6$, $p = .18$, all other F 's < 2.36 . Hence, there were no significant differences between the groups in the distance between ideal observer and utilization weights. A similar ANOVA for just the utilization weights showed no significant effects of group either (all F 's < 2.08). Thus, groups did not differ in the way in which utilization weights progressed over the trial sequence.

Another way to look at the correspondence between utilization and validity weights is by means of their correlation. Note that, if the response process of Equation 5 holds, this correlation is not affected by the response scaling parameter. For each participant and each card combination, we computed the correlation between utilization and ideal observer weights. These correlations are depicted in Figure 4, where they are plotted against overall performance (the score statistic of Equation 6). The mean correlation for card 1 was 0.44 ($SD = 0.37$) and the correlation ($df = 198$) was significant for 22 of the 25 participants. As Figure 4 shows, apart from three coefficients, the correlations were positive and many higher than .5. For card 2, the mean correlation was 0.18 ($SD = 0.37$) and the correlation was significant for 17 participants. Figure 4 shows that there are more negative correlations for this card than for card 1. The mean correlation for card 3 was 0.28 ($SD = 0.27$), with a significant coefficient for 19 participants. Like card 2, there was a relatively high number of negative correlations. The mean correlation for card 4 was 0.51 ($SD = 0.36$), with significance for all participants. In short, the correlation between utilization and validity weights were higher for the strongly predictive (card 1 and 4) than weakly predictive cards (cards 2 and 3). This is understandable, since accurately learning about these cards will boost categorization performance more than accurate learning of the weakly predictive cards⁶. A MANOVA, with the correlations for the four cards as dependent variables (after

⁶Indeed, Figure 4 shows that the utilization-ideal observer correlations for the strong cards were positively

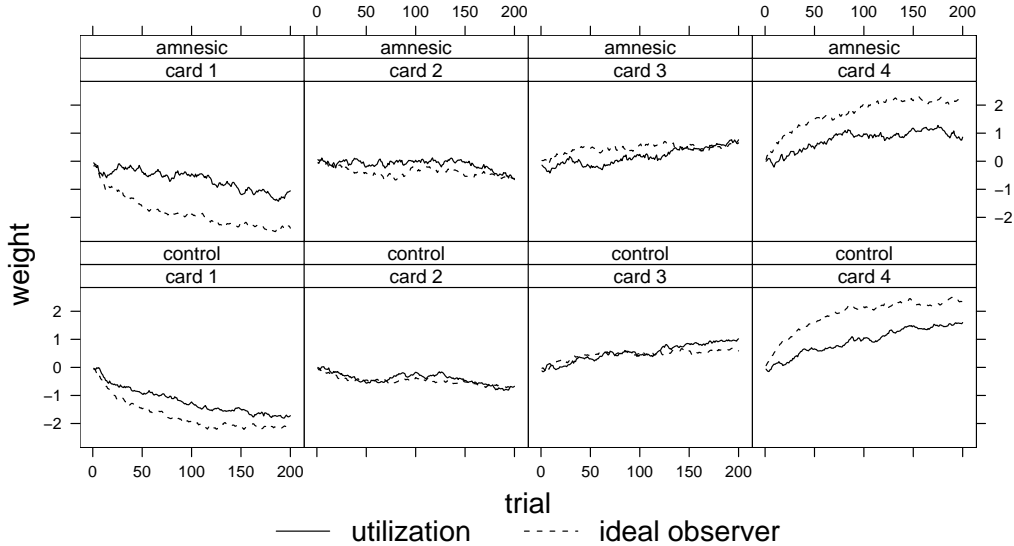


Figure 3. Average cue utilization weights (solid) and ‘ideal observer’ weights (broken) for the amnesic and control group, derived from the associative model.

Fisher’s Z transform), showed no reliable effect of group, Wilks’ $\Lambda = 0.85$, $F(4, 20) = .87$, $p = .50$.

In summary, the lens model analysis shows that most participants learned to use the cues appropriately. Especially for the strong cues, the cue utilization weights diverged quickly and in the expected direction. The correlation between utilization and ideal observer weights was stronger for the highly predictive cards than for the weakly predictive cards. Furthermore, a stronger utilization-ideal observer correlation boosted performance more for the strongly predictive than the weakly predictive cards. There was a negative relation between this correlation and performance for card 2. It may be that participants with a high correlation for this card utilized this card in lieu of the strongly predictive cards. Again, we see no clear differences between the amnesic and control group; not in cue utilization profiles (Figure 2), nor in the relation between cue utilization and ideal observer weights. While not a statistically reliable difference, the mean utilization profiles in Figure 3 do indicate that the under-utilization of the strong cards was more extensive for the amnesic than for the control group. This is consistent with the difference in learning rate reported earlier.

Task knowledge

After each block of fifty trials, participants rated the probability of the weather given each card. The average ratings for each block are depicted in Figure 5. A 1 between (group) \times 2 within (block, card) ANOVA showed a significant effect of card, $F(3, 69) = 13.97$,

related to performance, while this does not appear to be the case for the weakly predictive cards. We investigated this by computing rank correlation coefficients between the utilization-ideal observer weight correlation and performance. For cards 1 and 4, the rank correlation was significant and positive, $r = .74$, $p < .001$ and $r = .59$, $p < .01$, respectively. For card 2, the rank-correlation was significant but negative, $r = .44$, $p < .05$, while for card 3 it was non-significant, $r = .19$.

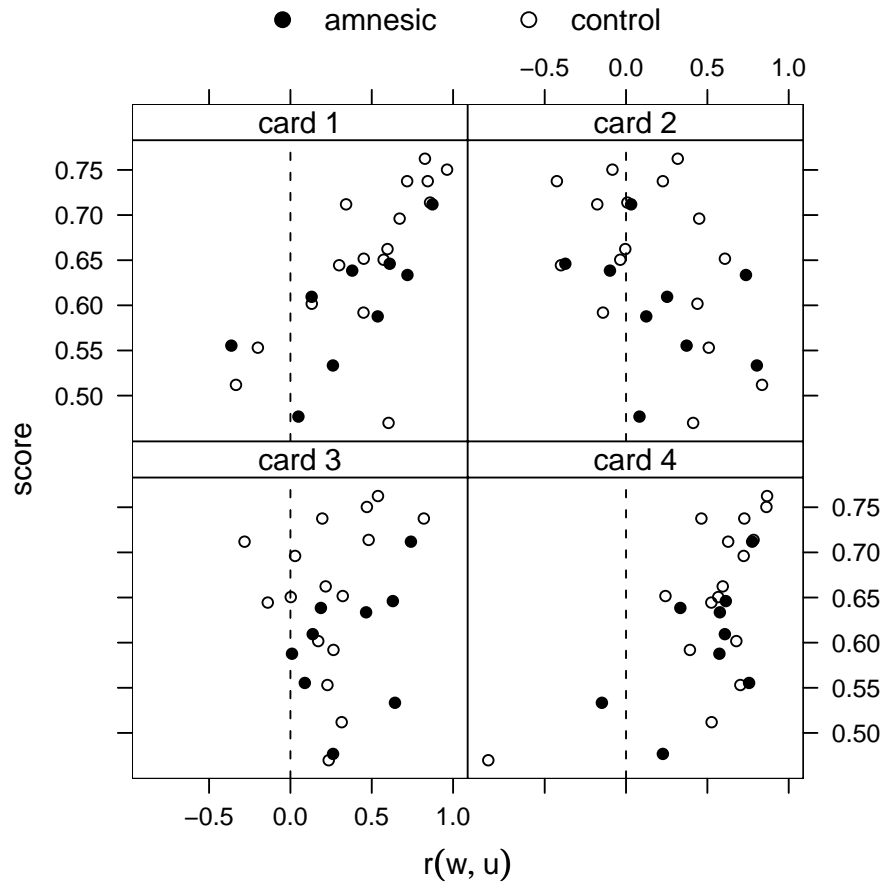


Figure 4. Plot of $r(w, u)$, the correlation between ideal observer (inferred cue validity) and cue utilization weights, against overall performance score.

MSe = .216, $p < .001$, a significant interaction between block and card, $F(9, 207) = 2.04$, MSe = .081, $p < .05$, as well as a significant interaction between block and group, $F(3, 69) = 3.02$, MSe = .081, $p = .036$. Other effects were not significant (all F 's < 1.90). The main effect of card shows that, overall, the judged probabilities were different for each card, with the means ordered in the expected direction (means are .31, .44, .55 and .74 for cards 1 to 4 respectively). The interaction between block and card indicates that participants improved their ability to discriminate between the probabilities of each card. Inspection of Figure 5 confirms this. By block 4, the average judged probabilities were reasonably close to the true probabilities (.2, .4, .6 and .8 for cards 1 to 4 respectively). The probabilities for card 1 and 2 were overestimated somewhat, as was the probability of card 3 by controls. The interaction between group and block appears to be a spurious effect, due to a higher overall estimated probability at block 2 by the control subjects. It is interesting to note that there was no interaction between group and card, $F(3, 69) = 1.39$, MSe = .216, $p = .25$, nor a significant three way interaction between group, card, and block, $F(9, 207) = 1.31$, MSe = .081, $p = .24$, indicating that there was no reliable difference between the groups

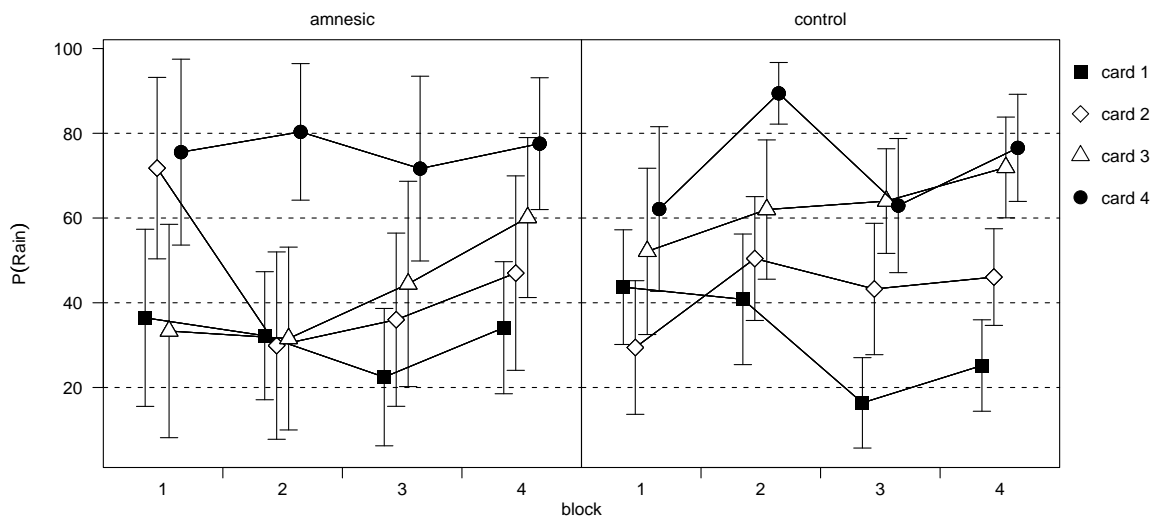


Figure 5. Averages and 95% confidence intervals of the probability ratings by block, for the amnesic and control group.

in task knowledge. The probability ratings in each group were highly correlated with the objective probabilities, amnesic group: $r(142) = .40$, $p < .001$, control group: $r(254) = .44$, $p < .001$.

While the previous analysis shows that the probability ratings conform to those of the environment, it does not consider what each subject could have known about the environment at the time of rating. Due to the probabilistic nature of the task, the environment as experienced by the participants may have differed from the theoretical environment, especially early on in the task. Hence, we should consider how well the ratings correspond to the probabilities they could have predicted based on their experience so far. To investigate this correspondence, we look at the relation between the ratings and the probabilities predicted by a normative model. This last value is derived from Equation 1 and the normative model weights as

$$P(\text{Rain}|\text{card}j) = [1 + \exp(-w_j)]^{-1},$$

where w_j is the inferred cue validity of card j . For each participant, we computed the correlation between the rated probabilities and these model derived probabilities (based on the weights at the trial of rating). These correlations ranged from 0.04 to 0.88 in the amnesic group ($M = 0.41$, $SD = .32$), and from -0.08 to 0.89 in the control group ($M = 0.44$, $SD = .31$); there was no significant difference between the groups, $t(22) = .20$, $p = .42$ (after Fisher Z -transform). Again, this shows that there is clear evidence for task knowledge, with no difference between the amnesic and control group. Finally, to investigate whether task knowledge was related to memory performance, we correlated the individual correlations computed above with the Wechsler memory scores listed in Table 1. These correlations ranged from .04 to .33, but were not significantly different from 0 (all p 's $> .12$). However, the individual correlations were strongly related to categorization performance, $r(22) = .75$, $p < .001$.

Self Insight

The previous analyses show that both amnesic and control participants had adequate knowledge about the task structure. To investigate whether the participants had insight into their response process, participants were asked to rate the importance of each card for their predictions after trial 50, 100, 150 and 200. Participants were quite variable in their use of the rating scale. Most participants only used the mid and end points of the scale, while some gave more precise ratings. As most participants appeared to use the rating scale as a three-point scale, we decided to treat it as such in the remaining analysis, categorizing ratings in the lower third of the scale as “low importance”, ratings in the middle third as “moderate importance” and ratings in the highest third as “high importance”. Our interest was in whether differences in importance ratings reflected differences in utilization. More precisely, we expected cues with high importance ratings to have a high absolute utilization weight. Figure 6 shows the average absolute utilization weight for each level of importance rating, separated by condition and block. To test the overall relation between utilization and perceived importance, we partitioned the total sum of squares of absolute utilization weights into a between participants, between blocks, and within blocks part, and tested for the effect of importance rating in each of these, focussing on a linear contrast for the three levels of importance rating⁷. To investigate group differences, we also tested for interactions between importance and group. Importance ratings accounted for a significant proportion of between participants differences in utilization, $F(1, 20) = 7.40$, $MSe = 1.96$, $p < .05$, indicating that participants who rated the cues as of low importance relative to other participants, utilized the cues less than those who rated the cues as of high importance. The interaction between importance and group was not significant, $F(1, 19) = 0.001$, $p = .98$. Importance ratings did not account for a significant proportion of between block differences in utilization, $F(1, 71) = 0.15$, $MSe = 0.47$, $p = .71$, nor was there a significant interaction with group, $F(1, 73) = 0.48$, $p = .49$. Hence, the block to block rise in utilization was not reflected by a rise in importance ratings. Finally, importance ratings accounted for a marginally significant proportion of within block variance in utilization, $F(1, 296) = 3.83$, $MSe = 0.48$, $p = .051$, indicating that cards with high importance ratings were utilized more than cards with low importance ratings. The interaction with group was again not significant, $F(1, 296) = 1.98$, $p = .16$.

While related, the importance ratings were rather noisy indicators of utilization, as can be seen in Figure 6. At the earlier blocks, importance ratings hardly discriminate between strong and weakly utilized cues. However, by block 4, the importance ratings do reflect the order of the absolute utilization weights, with cues with a high importance rating being utilized more than cues with a low importance rating. At block 4, the absolute utilization of low importance cues was significantly below the utilization of high importance cues; amnesic group: $t(13) = 2.10$, $p = .03$, control group: $t(32) = 1.96$, $p = .03$ (both one-sided). This was not the case in the other blocks. This suggests self-insight may have developed relatively late.

To summarize, the (marginally) significant relation between the rated importance of

⁷We also included a quadratic contrast in the model, since the relation between importance and utilization may not have been linear. The effect of this contrast was only significant between blocks, $F(1, 71) = 4.59$, $p < .05$, and showed no significant interactions with group, all F 's < 2.86 .

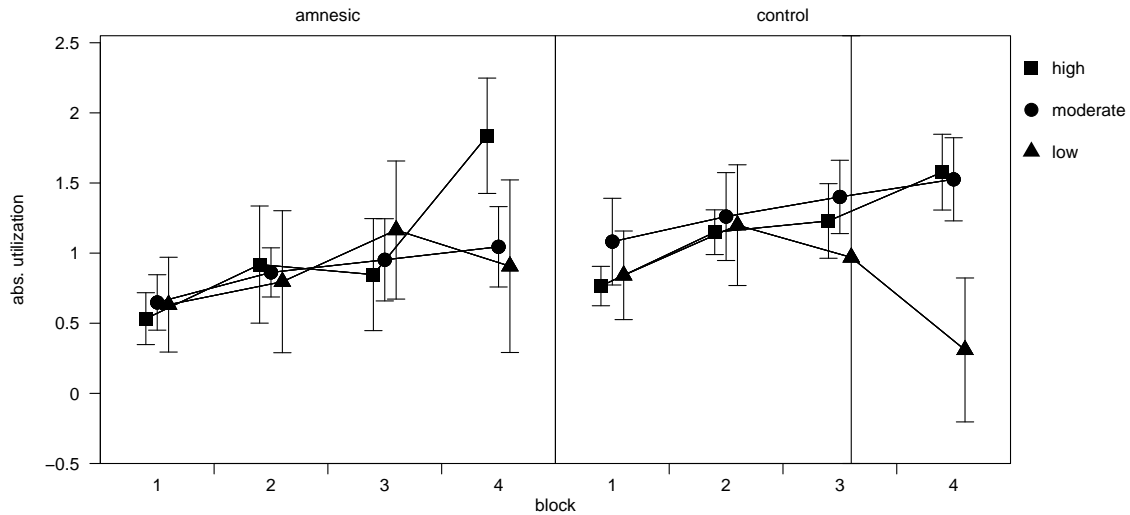


Figure 6. Averages and 95% confidence intervals of the absolute cue utilization weights by importance rating (low, moderate, high) and block, for the amnesic and control group. The confidence interval of one average (low importance, block 3, control group) was based on only two observations and very large (ranging from -2.35 to 4.29), and not plotted fully to increase discriminability of other averages.

cues and the strength of their utilization indicates that participants' had some insight into their response process, although this insight may have developed relatively late. Moreover, as the relation between importance ratings and utilization weights did not differ between the groups, we found no evidence for a relative impairment in individuals with amnesia.

Discussion

Previous research suggested that categorization performance is unimpaired in amnesia, while amnesic individuals lack declarative knowledge of task features. This has been taken as evidence for a multiple systems view of memory, in which categorization relies on implicit (procedural) memory, which is dissociated from explicit (declarative) memory. By adopting a lens model approach, we obtained a fine-grained analysis of the learning process in amnesic and control participants. Our results replicated the earlier finding of unimpaired category learning, but we found no evidence for impaired declarative knowledge in amnesia. Not only did both amnesic and control participants learn to use multiple cues to predict category membership, both groups developed accurate knowledge of the contingencies in the environment. Hence, the present research does not confirm the hypothesis of a dissociation between procedural and declarative memory in amnesia. Interestingly, we also found that categorization performance was significantly correlated with measures of immediate and delayed visual memory. So, explicit memory for visual material is strongly related to performance on a (visual) "implicit" task. This puts further doubt on the assumption (e.g. Knowlton et al., 1994) that the weather prediction task is procedural.

We showed that both implicit (categorization performance) and explicit knowledge (probability judgements for contingencies between cues and outcome) can result from a

single memory system, which sequentially forms associations between the cues and outcome. Not only did this associative model describe participants' category predictions, but we also found a strong relation between the parameters of this model and participants' probability judgements. In turn, these correlations were strongly related to categorization performance, supporting the single system expectation of a relation between categorization performance ("implicit" memory) and task knowledge ("explicit" memory). Since a single system account is both descriptively adequate and more parsimonious, we conclude that it is preferable over a multiple systems account of memory.

Implicit and explicit learning

Tasks such as the weather prediction task, requiring gradual learning and integration of probabilistic information, are often considered to involve implicit or procedural learning. Even when people perform well in these tasks, it is claimed that they have little or no insight into how they achieve this. This dissociation between performance and insight has not only been found for amnesic patients (e.g. Reber et al., 1996), but also for individuals without neurological problems (e.g. Evans, Clibbens, Cattani, Harris, & Dennis, 2003; Gluck et al., 2002). In contrast, Lagnado et al. (2006) found no evidence for a lack of task knowledge and self-insight. As in the present study, participants in their study showed accurate knowledge of the contingencies in the environment, and could distinguish between strongly and weakly utilized cues. What is the reason for the discrepancy of these latter results with those supporting the implicit/explicit distinction?

There are several reasons to question the validity of the results supporting the dissociation. First of all, the measures of explicit knowledge tend to be retrospective, while it is better to get multiple assessments as close as possible to the moment of judgement (Ericsson & Simon, 1984; Lovibond & Shanks, 2002). One possible caveat of this procedure is that posing explicit knowledge questions during the task may influence participants' later behaviour. More in particular, asking participants to judge cue-outcome contingencies may induce explicit knowledge which can be relied upon in later trials. However, comparison of groups who were posed explicit knowledge questions during or only at the end of the task (Lagnado et al., 2006) showed that this possibility is highly unlikely. Secondly, explicit tests of self-insight have relied on verbalization, which can hide self-insight which individuals find difficult to express verbally. While rule-based strategies can usually be accurately verbalized, this will be more difficult for strategies which rely on information integration (Ashby & Maddox, 2005). In such cases, one should pose questions in a format which resembles the accessible knowledge (i.e., the cue-outcome contingencies and the importance of the various cues for predictions). Finally, tests of explicit knowledge have been too vague, addressing task features irrelevant to performance. These methodological shortcomings were overcome in Lagnado et al. (2006) and the present study. In both cases, there was no evidence suggesting a lack of explicit knowledge accompanying probabilistic category learning. This puts severe doubt on the thesis that probabilistic category learning involves an implicit memory system inaccessible to conscious awareness.

While the probability ratings showed clear evidence of task knowledge, the importance ratings provided less convincing evidence for self-insight. Lagnado et al. (2006) reported good self-insight, but related the importance ratings to cue validity rather than utilization. The latter comparison should provide a stronger test, and the results indicated that self-

insight may have developed relatively late. This is understandable, as self-insight requires individuals to represent their response process, which is not directly related to the main task requirement of representing the cue-outcome contingencies. On the other hand, if learning was procedural, we would expect self-insight to be better than task knowledge. That we found the opposite indicates that learning was not mainly procedural.

Although these implications are interesting and worthy of further investigation, some caution is in place. That the importance ratings were rather noisy indicators of cue utilization may have been due to many reasons, including difficulty and individual differences in the use of the rating scale, and non-linear relations between cue utilization and importance ratings. As an indication of self-insight, we were trying to ascertain whether participants could discriminate between strongly and weakly utilized cues. However, the importance ratings did not require participants to compare the cues to each other. Pairwise comparisons, in which participants are asked on which of two cues they relied more, may be better measures for this purpose.

The lens model approach to learning

That we did not find qualitative differences in learning between amnesic and control participants is all the more surprising since our analyses were more sensitive to individual differences in learning than those in previous research. Previous research either focussed solely on the result of the learning process, modelled response strategies without regard for what each individual could have inferred about the environment given the information encountered so far, or considering individual differences in the relation between learned contingencies and predictions. The lens model approach overcomes these shortcomings. It provides a trial-by-trial overview of individuals' learning and response behaviour, and it allows for a direct comparison between what an individual could have known about the environment and how this knowledge is reflected in their predictions.

In the introduction, we argued for a distinction between learning and the response process. Where a learning process involves the gradual acquisition of knowledge of the contingencies in the environment, the response process involves the use of this knowledge as a basis of predicting the state of the environment. In the lens model approach, a distinction is made between inferred cue validity ('ideal observer') and cue utilization weights. The first concern trial-by-trial estimates of the relation between cues and outcome, while the latter concern trial-by-trial estimates of the relation between cues and response. As such, the inferred cue validity weights reflect individuals' knowledge states, while the cue utilization coefficients depend on these knowledge states and the response process.

If individuals learn as formalised in the learning model and respond by probability matching, the cue utilization and validity weights should be similar. Differences between the weights may have different sources. One possible source is a deviation in response scaling from probability matching. In our results, the absolute cue utilization weights were often lower than the absolute ideal observer weights. This reflects 'undershooting', where participants' predictions are more variable than the outcome. While undershooting may be solely the result of response scaling, it can also be due to a difference between the actual and ideal learning process. In our lens model analyses, we used fixed values for the learning rate parameters of the associative models. These values were chosen to result in good performance of the ideal observer model. However, there were clear indications

that the learning rate varied between individuals, and that the learning rate was usually lower than these fixed values. Hence, it is likely that difference between absolute cue utilization and ideal observer weights is due to participants' learning at a lower rate than the ideal observer. Using a similar lens model approach, Lagnado et al. (2006) found strong evidence for overshooting, while the control subjects in our study showed undershooting. It is likely that this discrepancy is due to a difference in population; participants in Lagnado et al.'s experiments were university students, while the control participants in our study were specifically matched to the amnesic group (who were older, had lower intelligence scores, etc., than the average university student, and can thus be expected to learn more slowly).

Regarding the learning rate, we found evidence that participants learned at a constant rather than decreasing rate. To our knowledge, this is the first research investigating this distinction. Previous implementations of associative learning models usually adopt the LMS learning rule, and hence assume a constant learning rate. The present result suggests that this assumption has been valid. We indicated that the Bayesian learning model implicitly incorporates a decreasing learning rate; later observations have relatively less effect on the parameter estimates than earlier ones. Hence, the better fit of the associative model to participants' responses may be partly due to this aspect of the model. To accurately capture individuals' learning process, Bayesian models may have to be adjusted to allow for a greater effect of later observations. In doing so, they could also account for the recency effects often found in human learning.

Learning in amnesia

We replicated the earlier findings of unimpaired early categorization performance in amnesia. However, in contrast to earlier studies, we found no evidence for impaired performance of amnesic individuals later in the task. This lack of discrimination may be due to several causes. First of all, the amnesic group in this study may have consisted of particularly good learners. Secondly, the control participants in the study may have been rather poor learners. The final explanation is that, due to the small number of amnesic participants, the test employed may have lacked statistical power. The average percentage of optimal responses of the amnesic and control group (see Footnote 4) were similar to those in Knowlton et al. (1994) and Knowlton et al. (1996), who used a version of the weather prediction task with different contingencies. Hopkins et al. (2004, Experiment 1), used the same contingencies as the present experiment; over 200 trials, their control participants averaged 83.8% ($SD = 8.3$) optimal responses, compared to 66.8% ($SD = 11.6$) in our control group, which is a significant difference, $t(23) = 3.76$, $p < .001$. The difference in the performance of the amnesic groups is negligible, 61.6% ($SD = 10.6$) vs 60.3% ($SD = 9.9$) in Hopkins et al.'s and the present study respectively, $t(23) = 0.30$, $p = .38$. While this indicates that the control participants in our study may have been particularly poor learners, we should note that there are significant differences in age and education level between the control groups of both studies. Moreover, we found a significant increase in performance over blocks, while this was not evident in Hopkins et al.. As such, there is no question that our control group learned to categorize the stimuli, although perhaps at a lower rate than others.

Although categorization performance of our control group was lower than that of healthy participants in other studies, their performance was similar to that of the controls

in Knowlton et al. (1994), who did find a reliable difference with a slightly different version of the task. As such, the lack of a significant difference in the present study may have been due to lack of statistical power. The overall better performance of the controls is consistent with this explanation. The small number of amnesic participants is not particular to our study. For instance, the experiments of Eldridge et al. (2002), Knowlton et al. (1994), and Reber et al. (1996) all had eight amnesic participants, while Hopkins et al. (2004) had nine. Lack of statistical power is common in neuropsychological studies that rely on patient populations (Zaki, 2004). Since our argument partly rests on a lack of difference between amnesic and control participants, power may pose a more serious problem here than elsewhere. However, we should be clear that rather than predicting no difference between amnesic and unimpaired individuals, a single system account predicts a relation between categorization performance and task knowledge. As both rely on the same memory system, an impairment should affect both. Although we cannot rule out the possibility that lack of statistical power resulted in the absence of reliable differences in categorization performance, task knowledge, and self-insight, correspondence between the results is consistent with the single system account. Moreover, we found reliable evidence for task knowledge in the amnesic group, and this task knowledge was reliably related to categorization performance. Both these results are not readily predicted by a multiple systems account.

While we found no qualitative differences between amnesic and control participants in the way in which they learn, we did find evidence for a quantitative difference between the groups, in that learning appeared to be slower in the amnesic group than in the control group. Allowing for individual differences in the rate of associative learning, we found that control participants showed a higher learning rate than amnesic participants. A similar difference was found for the Bayesian learning model, where the response scaling parameter was lower in the amnesic than control group. According to this model, amnesic individuals showed less reliance on the learned contingencies than controls. While the interpretation of the group differences depends on the model chosen, both show that the amnesic group tended to ‘undershoot’ more than the control group. No matter what the preferred explanation, the difference between groups in learning rate or response scaling would result in a slower increase in performance for the amnesic as compared to the control group. Hence, the difference in later categorization performance between normal and amnesic participants found in earlier studies can be explained in a single system account. There is no need to posit a later reliance on declarative memory for control participants to explain such divergence. Since both amnesic and control participants showed accurate knowledge of the contingencies in the environment, both could have developed such strategies.

Since we have found no qualitative difference in the way in which amnesic and control participants learn, nor any differential access to what they have learned, it may be that amnesia mainly affects the rate of learning in probabilistic categorization tasks. Whether or not one finds differences between control and amnesic groups will then depend on the size of the difference in learning rate and where in training one looks for such a difference. At the start of training, the chance of finding significant differences will be relatively low; only very large differences in learning rate will result in enough power to distinguish groups through performance. In later training stages, the difference in learning rate will result in more marked differences in performance. However, this difference should diminish again after prolonged training, since learning rate (within bounds) does not affect asymptotic

performance. If learning rate is the main cause of performance differences, we would expect the difference between control and amnesic groups to disappear asymptotically. Given enough practice, just like control subjects, amnesic individuals should be able to reach an optimal level of performance.

It may appear we suggest no difference between amnesic and normal individuals in memory function. However, this does not reflect our position. Rather, we conclude that in tasks such as the weather prediction task, there are no clear differences in the ways in which individuals with and without amnesia learn about cue-outcome contingencies. Both groups learn to use the cues appropriately to form predictions. We do not propose that amnesic individuals are not affected by severe memory problems; the clinical tests show otherwise. Individuals with amnesia may lack declarative knowledge of certain task features, or even the whole learning episode. However, these forms of declarative memory are inconsequential to the task. Knowing exactly how many squares a card contains will not increase the accuracy of predictions. Nor will the recollection of specific instances of cue-outcome pairs; remembering the squares and diamonds cards were paired with sun on trial 21 does not directly aid performance, but remembering that the squares and diamonds cards were more often paired with sun than rain does. Amnesic patients may be impaired in remembering specific instances of the training episode. Such episodic memories require binding events to the context in which they occur, and the medial temporal lobe system has been taken to be particularly involved in such contextual or relational memory (Cohen et al., 1999; O'Reilly & Rudy, 2001). In this view, the brain regions affected in amnesia aid the recollection of unique events by binding together various components to form unique representations. Probabilistic classification is relatively spared because it does not require such contextual binding, but rather the extraction of repetitive patterns of co-occurrence. A process which incrementally learns cue-outcome associations will suffice for this purpose. Commonly, such a process is assumed to be implicit. However, participants in our experiment showed accurate knowledge of the cue-outcome contingencies, as well as insight into the importance of cues for their predictions. Hence, we found no dissociation between implicit learning and explicit awareness. While learning may not rely on explicit recollection of observed cue-outcome patterns, to classify the learning process as implicit, in the sense that its elements are inaccessible to conscious awareness, is mistaken.

References

- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.
- Bishop, D. V. M. (1989). *Test for the reception of grammar: Manual*. Oxford, UK.
- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition (Special issue). *Trends in Cognitive Sciences*, *10*(7).
- Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., & Nash, C. (1999). Hippocampal system

- and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus*, *9*, 83–98.
- Cooksey, R. W. (1996). The methodology of social judgement theory. *Thinking and Reasoning*, *2*, 141–173.
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system test*. San Antonio, TX: The Psychological Corporation.
- Eldridge, L. L., Masterman, D., & Knowlton, B. J. (2002). Intact implicit habit learning in Alzheimer's disease. *Behavioral Neuroscience*, *116*, 722–726.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- Evans, J. S. B. T., Clibbens, J., Cattani, A., Harris, A., & Dennis, I. (2003). Explicit and implicit processes in multicue judgment. *Memory & Cognition*, *31*, 608–618.
- Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences of the USA*, *103*, 11778–11783.
- Friedman, D., & Massaro, D. W. (1998). Understanding variability in binary and continuous choice. *Psychonomic Bulletin & Review*, *5*, 370–389.
- Friedman, D., Massaro, D. W., Kitzis, S. N., & Cohen, M. M. (1995). A comparison of learning models. *Journal of Mathematical Psychology*, *39*, 164–178.
- Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, *49*, 87–115.
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227–247.
- Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the “weather prediction task”? Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*, 408–418.
- Goldstein, W. M. (2004). Social judgment theory: Applying and extending Brunswik's probabilistic functionalism. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making*. London: Blackwell.
- Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia*, *42*, 524–535.
- Kelley, H., & Friedman, D. (2002). Learning to forecast price. *Economic Inquiry*, *40*, 556–573.
- Kinder, A., & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience*, *13*, 648–669.
- Kitzis, S. N., Kelley, H., Berg, E., Massaro, D. W., & Friedman, D. (1998). Broadening the test of learning models. *Journal of Mathematical Psychology*, *42*, 327–355.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.
- Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*, 106–120.
- Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple cue learning. *Journal of Experimental Psychology: General*, *135*, 162–183.
- LeCun, Y., Bottou, L., Orr, G. B., & Miller, K. R. (1998). Efficient backprop. In G. B. Orr & K. R. Miller (Eds.), *Neural networks: Tricks of the trade* (pp. 9–50). Berlin: Springer-Verlag.
- Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 3–26.
- Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225–252.
- McClelland, J. L., & Rumelhart, D. E. (1986). Amnesia and distributed memory. In D. E. Rumel-

- hart, J. L. McClelland, & The PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 503–528). Cambridge, MA: MIT Press.
- Meeter, M., Myers, C. E., Shohamy, D., Hopkins, R. O., & Gluck, M. A. (2006). Strategies in probabilistic categorization: results from a new way of analyzing performance. *Learning & Memory, 13*(2), 230–239.
- Newell, B. R., Lagnado, D. A., & Shanks, D. R. (in press). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review*.
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247–255.
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review, 108*, 311–345.
- Poldrack, R. A., Clark, J., Par-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546–550.
- Poldrack, R. A., & Rodriguez, P. (2004). How do memory systems interact? Evidence from human classification learning. *Neurobiology of Learning and Memory, 82*, 324–332.
- Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience, 110*, 861–871.
- Robbins, H., & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics, 22*, 400–407.
- Sage, J. R., Anagnostaras, S. G., Mitchell, S., Bronstein, J. M., Salles, A. D., Masterman, D., et al. (2003). Analysis of probabilistic classification learning in patients with Parkinson's disease before and after pallidotomy surgery. *Learning & Memory, 10*, 226–236.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of statistics, 6*, 461–464.
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making, 15*, 233–250.
- Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review, 94*, 439–454.
- Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*, 676–686.
- Wechsler, D. (1997). *Wechsler memory scale: Third edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio, TX: The Psychological Corporation.
- Zaki, S. R. (2004). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin & Review, 11*, 1048–1054.
- Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1160–1173.

Appendix

Relation between utilization \mathbf{u} and inferred cue validity \mathbf{w}

In this appendix, we show the effect of the response scaling process on the cue utilization weights. We begin by noting that for a logistic regression model, the log-odds of a

response is a linear function of the cue values

$$\log \left(\frac{P(r = 1|\mathbf{x})}{P(r = 0|\mathbf{x})} \right) = \mathbf{u}^\top \mathbf{x}.$$

When

$$P(r = 1|\mathbf{x}) = \frac{P(y = 1|\mathbf{x})^\lambda}{P(y = 1|\mathbf{x})^\lambda + P(y = 0|\mathbf{x})^\lambda}$$

as in Equation 5, we can write the log-odds of a response as a function of the log-odds of the outcome

$$\log \left(\frac{P(r = 1|\mathbf{x})}{P(r = 0|\mathbf{x})} \right) = \lambda \log \left(\frac{P(y = 1|\mathbf{x})}{P(y = 0|\mathbf{x})} \right).$$

Since

$$\log \left(\frac{P(y = 1|\mathbf{x})}{P(y = 0|\mathbf{x})} \right) = \mathbf{w}^\top \mathbf{x},$$

we have

$$\mathbf{u} = \lambda \mathbf{w}.$$