Review

Probabilistic categorization: How do normal participants and amnesic patients do it?

M. Meeter[a],[*], G. Radics[a], C.E. Myers[b], M.A. Gluck[c], R.O. Hopkins[d],[e]

[a]Department of Cognitive Psychology, Vrije Universiteit Amsterdam, The Netherlands
[b]Department of Psychology, Rutgers University, NJ, USA
[c]Center for Molecular and Behavioral Neuroscience, Rutgers University, NJ, USA
[d]Psychology Department and Neuroscience Center, Brigham Young University, UT, USA
[e]Department of Medicine, Pulmonary and Critical Care Division, LDS Hospital, UT, USA

Abstract

In probabilistic categorization tasks, various cues are probabilistically (but not perfectly) predictive of class membership. This means that a given combination of cues sometimes belongs to one class and sometimes to another. It is not yet clear how categorizers approach such tasks. Here, we review evidence in favor of two alternative conceptualizations of learning in probabilistic categorization: as rule-based learning, or as incremental learning. Each conceptualization forms the basis of a way of analyzing performance: strategy analysis assumes rule-based learning, while rolling regression analysis assumes incremental learning. Here, we contrasted the ability of each to predict performance of normal categorizers. Both turned out to predict responses about equally well. We then reviewed performance of patients with damage to regions deemed important for either rule-based or incremental learning. Evidence was again about equally compatible with either alternative conceptualization of learning, although neither predicted an involvement of the medial temporal lobe. We suggest that a new way of conceptualizing probabilistic categorization might be fruitful, in which the medial temporal lobe help set up representations that are then used by other regions to assign patterns to categories.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Probabilistic categorization; Basal ganglia; Prefrontal cortex; Hippocampus

Contents

---

*Corresponding author. Tel.: +31 20 4448993.
E-mail address: m@meeter.nl (M. Meeter).

1. Introduction

In probabilistic category-learning tasks, various cues are probabilistically (but not perfectly) predictive of class membership. They have been used extensively in cognitive and neuropsychological research, especially because they are thought to provide insight into implicit forms of learning, cognitive flexibility and the use of feedback signals in the brain. These tasks have also been used to elucidate cognitive deficits in several patient populations, including patients with medial temporal lobe damage and patients with Parkinson's disease (Knowlton et al., 1994, 1996; Hopkins et al., 2004; Shohamy et al., 2004).

While probabilistic categorization has been widely embraced in cognitive neuroscience research, it is still unknown exactly how individuals solve such tasks. It could be that participants attempt to find a rule underlying the category assignments, but repeated exposure to exemplars could also slowly lead to a tendency for subjects to group similar stimuli in the same categories. A third way to solve the task is that subjects could simply memorize an answer for each individual cue combination. Thus, there are several ways in which a subject could approach probabilistic categorization tasks and achieve significantly better-than-chance performance. Especially relevant from a theoretical viewpoint is whether probabilistic categorization can be considered a procedural memory task, or is in part or whole a declarative memory task. Probabilistic category learning was developed to tap procedural memory (Knowlton et al., 1996), but more recently researchers have been arguing for a declarative component to the task (Meeter et al., 2006), or even that the task is entirely declarative (Lagnado et al., 2006). Here, we will discuss evidence for such a reclassification of probabilistic categorization as a declarative memory task. As with similar debates, the hardest questions to answer are what it means for the task to be procedural, and what would count as evidence for a reliance of the task on procedural or declarative memory. We will first turn to these questions.

2. What does it mean for a task to be procedural?

Procedural memory refers to ''knowing how'' to do things, and usually implies a lack of conscious access to the memories underlying performance. Procedural memories are acquired slowly through training, and are highly specific: they underlie a precise skill or procedure. Declarative memory refers to ''knowing that'', and underlies our knowledge of facts and events (Cohen and Squire, 1980). Declarative memories are flexible, in that they can underlie multiple kinds of performance. They can be reflected upon consciously, are representational (i.e., represent something in the world), and seem to rely on the medial temporal lobe (Squire, 2004). A similar distinction is that made between implicit and explicit memories (Graf and Schacter, 1985); memories are explicit when they can be accessed consciously, while implicit memories are those that influence behavior without conscious awareness. Although procedural and implicit seem to denote the same construct, there is a difference in focus that may be significant in some contexts. Indeed, a recent model of skill learning firmly disconnects the two, arguing for example for explicit procedural memories (Sun et al., 2005).

These distinctions map only imperfectly to a typology of learning in deterministic categorization, different from probabilistic categorization in that each pattern is always associated with a particular outcome. Say that a participant must classify colored line segments as belonging to either category X or category Y. If there is some easily verbalizable rule underlying the classification, such as that all red items belong to X and all blue items to Y, participants will usually discover the rule and use it. Ashby and Ell (2001) call such tasks rule-based category-learning tasks. The underlying model of learning is hypothesis testing, in which rules are discarded when they lead to wrong responses. Rule-based learning could be seen as either declarative, as the rules are presumably flexible and representational, or procedural, as the rules describe how to do a task (for Sun et al., 2005, rule-based learning is explicit and procedural).

In other tasks, there may be no easy rule: for example, long bluish segments belong to X, but not if they are too long or too blue. In such cases, participants have to integrate information from multiple dimensions, and slowly learn where in the multidimensional stimulus space the category boundaries are. Ashby and Ell (2001) term this an ''information integration'' approach, but we will here talk of incremental learning. This is because the integration of information across dimensions seems less essential than the slow, incremental learning of category boundaries. Such learning is sometimes also seen in tasks where integrating information across dimensions is not important, for example in children (Raijmakers et al., 2001). Incremental learning seems to fit characterizations of procedural learning in that it is slow, but is not necessarily inflexible given the variance over time often seen in categorization performance.

A special case of tasks in which no rule can be found is when there are so few exemplars that participants try to memorize all exemplars, and just remember what category each exemplar belonged to.

These kinds of learning are generally linked to different brain systems (Ashby and Ell, 2001). Tasks that allow learning through rules appear to rely on the basal ganglia and the prefrontal cortex. Incremental learning tasks are thought to rely most on the basal ganglia. Finally, tasks involving memorization of exemplars are affected mostly by medial temporal lobe lesions. Since the medial temporal lobe is usually thought to be the substrate of declarative memory, such learning would most likely be declarative.

2.1. Application to probabilistic categorization

Can these distinctions be mapped onto the ways in which individuals could solve probabilistic categorization tasks? One such task is the weather prediction task, or WP task. In this task, participants are shown sets of "tarot cards" that might predict the weather. Four cards can be part of these sets, each linked to four roles. The four cards are usually referred to as cues, while the sets of cards are referred to as patterns. One cue strongly predicts rainy weather (typically with 80% likelihood), one cue weakly predicts rain (typically with 60% likelihood), one cue strongly predicts sunny weather, and one weakly predicts sun (Fig. 1). At each trial, participants are shown a pattern consisting of one to three of these cues, and have to indicate whether rainy or sunny weather is more likely. The participants are given feedback. Participants are not instructed regarding the predictive association of the different cues, but have to learn via trial and error which patterns predict rainy weather, and which predict sunny weather.

Rule learning as the underlying mechanism would suggest that participants go about this task by attempting to find a rule to categorize all patterns (e.g., if the card with squares is present, then respond "sun", else respond "rain"). Strict testing of possible rules would be impractical

in probabilistic categorization, as no rule will lead to 100% correct performance. Instead, participants may try new rules of sets of rules whenever they feel too many responses were wrong in close temporal proximity. Incremental learning would suggest that subjects learn slowly, over trials, how to respond to individual cues (e.g., they could slowly discover that patterns including the squares card are more likely to be followed by sunny weather). Memorization would be akin to that participants learn to respond in a certain way to each individual pattern (e.g., the pattern with both squares and circles cards present usually means "sun").

None of these alternatives have been proposed in quite these words. Several papers have suggested that probabilistic categorization may rely on a habit-learning system encompassing the basal ganglia, and a declarative memory system centered on the medial temporal lobe (Poldrack et al., 2001; Moody et al., 2004; Foerde et al., 2006). In these papers, the habit-learning system is described in ways that suggest incremental learning. The declarative learning is not described in much detail, and might encompass either rule-based or memorization learning.

Lagnado et al. (2006) provided a concrete proposal for how learning in probabilistic categorization might be incremental. They argued that learning in the weather prediction task is gradual, and based on incrementally learning the extent to which each of the four cards predicts either sun or rain (i.e., finding the *weight* of each card). Such learning cannot be characterized as either implicit or procedural. Lagnado et al. (2006) argued that participants are aware of their cue weights: When asked to judge the contribution of cards to category assignment, subjects made explicit judgments that resemble the card weights revealed through their performance. Moreover, weights were not inflexible, one of the main characteristics of the outcomes of procedural learning. Data from individual participants showed, instead, that weights changed substantially throughout the experiment. Meeter et al. (2006) provided a conceptualization of learning that is suggestive of rule-based learning. They suggested that participants base their responses in probabilistic categorization tasks on a *strategy*, a rule or set of rules that assigns patterns to the two categories of sun and rain. An example of such a strategy is the "strong rain single-card strategy". This strategy assigns all patterns containing the strong rain card (the squares card in Fig. 1) to the "rain" category, and all other patterns to "sun". Fourteen strategies were defined, all of which led to above-chance performance and were simple enough to be formulated and used. Meeter et al. (2006) identified discrete shifts in the strategy used by participants over the course of an experiment, termed strategy shifts. The strategy shifts are akin to shifting from one rule to another, as occurs in rule-based learning. Meeter et al. (2006) did not commit to explicitly formulated rules, arguing only against the underlying learning being incremental. The third alternative, that participants memorize pattern–response
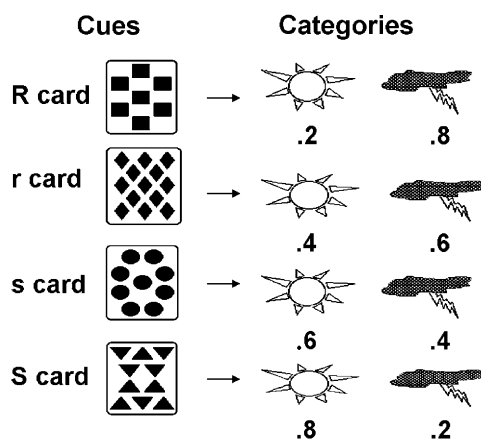

Fig. 1. Four cards in the weather task, and the likelihoods with which they predict the outcomes, rain and sun. The strong rain ("R") and sun ("S") cards each predict the weather (rain or sun) with 80% probability, while the weaker rain ("r") and sun ("s") cards each predict the outcome with 60% probability. One, two or three cards are presented on each trial, and the probability of each outcome on a given trial is a function of the probabilities of all cards present on that trial.

combinations, has not been put forward, but will be returned to in Discussion section.

Can a choice between these alternatives be made on empirical grounds? The positions appear to make contrary predictions in at least two areas. A first set of contradictory predictions concerns the trajectory of learning. Incremental learning approaches predict gradual learning, as exemplified by gradient descent algorithms. Rule-based learning, on the other hand, predicts a ''jerky'' learning trajectory, with sudden jumps in performance. Moreover, the two approaches suggest that different techniques are best suited to predict performance of participants: one geared towards uncovering weights of cards and the other geared towards uncovering strategies.

Second, the positions predict different brain substrates for probabilistic category learning. If probabilistic categorization is primarily or partly an incremental learning task, deficits should be apparent in patients with basal ganglia abnormalities. If, on the other hand, probabilistic categorization is primarily or partly a rule-based learning task, deficits should occur both in patients with basal ganglia and in patients with prefrontal lobe lesions. Neither position would prima facie predict deficits in patients with medial temporal lobe damage, which would, however, be expected if memorization plays a role.

In the following text, we will first discuss how normal performance in the weather prediction task can best be analyzed, followed by evidence from patient populations with damage in neural areas thought to be involved in probabilistic categorization.

3. What can we learn from a precise look at performance?

Traditionally, category-learning data is analyzed by calculating the proportion of optimal responses over the course of the experiment. The resultant learning curves depict *how fast* participants solve the task, but do not divulge much about *how* participants solve it. Two richer ways of analyzing performance have been introduced. One is strategy analysis, introduced by Gluck et al. (2002) and fine-tuned by Meeter et al. (2006); the other is rolling regression, adapted to the weather prediction task by Lagnado et al. (2006) following Kelley and Friedman (2002). We will discuss both in turn.

3.1. Strategy analysis

Gluck et al. (2002) recognized that in weather prediction, responses of participants to particular stimuli may occur in consistent patterns that are informative about how participants approach the task; they termed these consistent patterns *strategies*. Gluck et al. considered a finite set of four basic strategies, and were able to show that the behavior of most healthy participants was consistent with one of these strategies. Gluck et al. were also able to show a progression in individual participants from simple strategies to more complex ones as the experiment progressed.

Meeter et al. (2006) later extended and elaborated on the strategy analysis, and it is this extension that we will discuss here.

On any one trial in weather prediction, participants are presented with a pattern consisting of one, two, or three cards, and have to give a binary response (i.e., ''sun'' or ''rain''). This response can be thought of as being based on a disposition to answer ''sun'' or ''rain'' to the presented pattern with a certain probability. In fact, the participant might have such a disposition to each of the patterns in the task. Strategy analysis is an attempt to infer this set of dispositions for each participant from series of trials. From one response on a trial, it is impossible to deduce the set of dispositions of the participant towards all patterns. Even the disposition to the specific pattern presented on a trial cannot be deduced (e.g., if the participant answered ''sun'', only a likelihood of 0 of answering ''sun'' is ruled out; all other values are still possible). However, if the set of dispositions of the participant remains constant over a number of trials, it might be possible to identify the disposition from the pattern of responses.

Meeter et al. (2006) showed, using Monte Carlo simulations, that such identification of dispositions was possible, provided that the set of possible dispositions is limited. Possible sets of dispositions are called strategies in strategy analysis, and their number is limited by two principles: that all strategies lead to above-chance performance, and that they must be simple enough to be formulated and used. Each strategy is formulated as an ideal type of responses, which is fitted to series of 24 trials. In this way, the strategies used by a participant can be monitored throughout the experiment, and a switch from one strategy to the next can be pinpointed with reasonable precision.

Strategy analysis makes two strong assumptions. First, that there are stable states in performance that map onto the set of strategies and, second, that learning involves rather discrete switches that typically involve all patterns at once. The second assumption is the one that ties strategy analysis to rule-based learning: if responses to each cue evolve independently of one another, then learning will typically not affect all patterns at once; thus it will not be equivalent to the kind of switches strategy analysis looks for. Moreover, if dispositions of participants towards cues or patterns evolve slowly over trials, then any discrete switch in performance is an artifact.

3.2. Rolling regression

An alternate way to describe subject learning in the weather prediction task is rolling regression analysis (Lagnado et al., 2006). Rolling regression analysis is based on a model of learning in which participants try to uncover the weights of each card, and then combine the weights of the cards in a pattern to compute the odds of responding ''sun'' or ''rain'' to that pattern. The weights quantify how much each card predicts rain or sun, in more or less the

same way that regression coefficients quantify the relation between independent and dependent variables. Rolling regression analysis tries to uncover the weights through a logistic regression of participants' responses. If, for example, a participant tends to respond "rainy" to patterns that contain the square card, the analysis will lead to a high weight for the squares card (weights are positive for those cards that predict rain, and negative for those that predict sun). Lagnado et al. (2006) showed that when data are analyzed in this fashion, the weights of cards move away from 0, for most participants during the experiment, to positive values for cards that predicted rain, and negative values for cards that predicted sun.

Just as strategy analysis assumes a certain form of learning on the part of the participants, the same is true for rolling regression analyses. Rolling regression analysis assumes that learning is incremental, and consists of slowly developing weights for the four cards. The weights are then integrated into odds of responding "rain" or "sun" for a particular pattern of cards. It is fair to say that both strategy analysis and rolling regression analysis make strong assumptions about the ways in which participants are learning, and that the analyses are valid only if participants are indeed learning in the assumed manner. This begs the question of which manner of learning is in fact more often used by humans learning the weather prediction task. One way to investigate which form of learning is more prominent is to investigate which type of analysis is best at extracting patterns from human performance in probabilistic categorization tasks.

4. Predicting responses of healthy subjects

To test which type of analysis is superior, we applied analyses to an existing set of data collected from healthy young adults (university students) performing the weather prediction task. We set out to ascertain how well the analyses could predict an individual's future responses, on the basis of that individual's prior responses. If the analysis does not predict subsequent responses, it would indicate that the analysis is not a good descriptor of the response pattern—at least for that individual.

We analyzed data first reported by Gluck et al. (2002), Experiment 2. In brief, participants were 30 Rutgers University undergraduates (17 female, mean age 20.7 years) receiving class credit for their participation. They were given a 200-trial weather prediction task; on each trial, they were given 1 of 14 patterns consisting of the cards shown in Fig. 1, and asked to predict whether the weather would be rain or sun. After their response, participants received visual feedback about the actual weather outcome. Two hundred trials were generated to satisfy the card-outcome probabilities shown in Fig. 1; ordering of the trials was random but fixed across subjects. Responses were scored as "optimal" based on whether participants predicted the weather outcome most often associated with the current pattern, independent of the
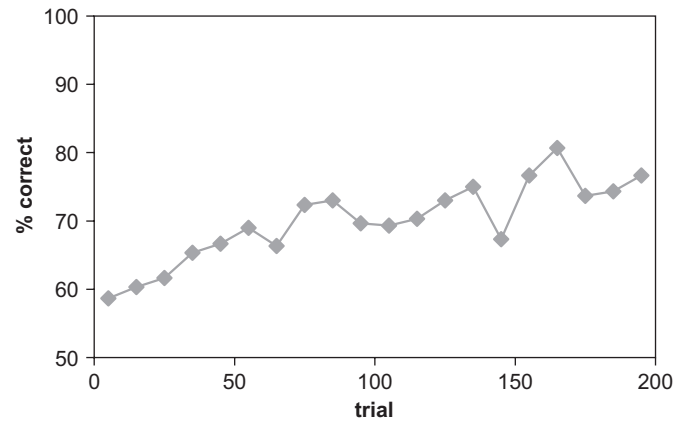

Fig. 2. Learning curve for young adults in the study of Gluck et al. (2002) for the weather task.

actual weather on the trial. Fig. 2 shows performance of participants over the 200 trials, as reported in Gluck et al. (2002). From a level close to chance, performance increased to about 80% optimal answers.

4.1. Fitting responses

Fig. 3 shows the outcomes of the analyses, both for all participants and for three individual subjects. The strategy analysis yields a progression from simple to complex strategies (also see Meeter et al., 2006). Early in training, most participants were best fit by simple strategies, with few or none best fit by a strategy in which all responses are optimal. By the end of training, more than half of participants were best fit by either an optimal strategy or by a strategy of intermediate complexity (Fig. 3A). At the individual level, however, progress was not smooth. While some participants progressed relatively smoothly from simple to complex strategies (e.g., participant 1 in Fig. 3B), others switched back from simple to no strategy (e.g., participant 2) or from complex to simpler strategies (e.g., participant 3 and participant 1 at the end of the experiment, see Fig. 3B).

The outcomes of the rolling regression analysis are shown in Fig. 4A. Weights of the cards were capped at 10 before averaging, because some reached very high values that distorted the results. Participants on average gave negative weights to cards that predict rain, and positive weights to cards that predict sun. Moreover, the weights of the "strong" cards (i.e., the cards with strong predictive power) were larger than those of the "weak" cards. Replicating Lagnado et al. (2006), card weights exceeded the values that corresponded to the cards' objective predictive power (gray lines in Fig. 4A). Learning seemed to be gradual, with card weights moving away from 0 throughout the experiment.

This was again not what appeared at the individual level. Fig. 4B–D shows the weights computed from the performance of the same three participants of whom fitted strategies are shown in Fig. 3B. For participants 1 and 3
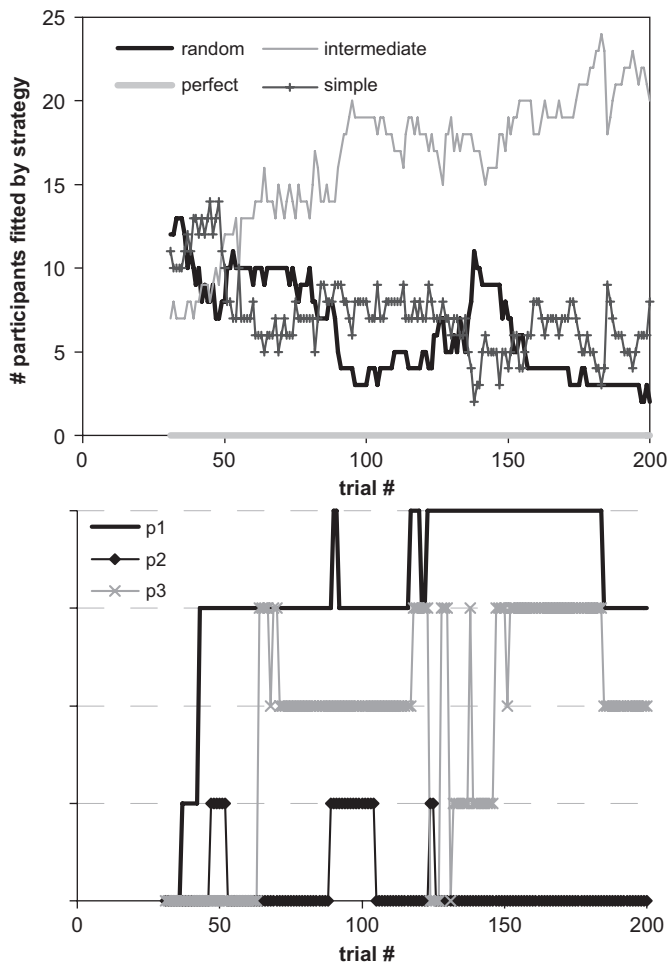
Fig. 3. Outcomes of the strategy analysis. (A) Number of participants best fit, at different trials, by either no strategy ("random"), a simple strategy ("simple"), a strategy of intermediate complexity ("intermediate"), or a strategy that approaches the true category structure ("perfect"). "Simple" strategies consist of singleton strategies in which participants respond correctly to individual card patterns but guess on patterns with more than one card, and single-cue strategies in which the participant bases the response on the presence or absence of one of the four cards. (B) Learning trajectory of three participants (p1, p2 and p3) throughout the experiment. For each participant, the line gives the strategy that best fits a participant at the trials 31–200 (no strategy can be fitted to the first 30 trials). Strategy groupings as in part (A), except that singleton and single-cue strategies are now split out.

(Fig. 4B and D), weights brusquely changed from low to high values. The trials at which this occurred were generally those at which the strategy analysis identified a switch from one strategy to the next. For participant 2, who was best fit by the "random" strategy throughout most of the experiment, weights changed smoothly but remained close to zero.

4.2. Response prediction

The results of both analyses were used to generate predictions for the responses. For strategy analysis, the strategy fit on trials $t-d$ to $t-1$ was used to generate a

prediction for trial t. For example, if a strong rain single-cue strategy (respond "rain" of strong rain card is present in the pattern, else respond "sun") fit best on a participant's responses on the d previous trials, the prediction for trial t would be "rain" if the strong rain card was present in the pattern, and else it would be "sun". For rolling regression analysis, the weights fit on trials $t-d$ to $t-1$ were used to generate the prediction. Weights were combined into an odd of a rain response, which was then transformed into a likelihood. Likelihoods above 0.50 were taken to predict a rain response, those below 0.50 as predicting a "sun" response. Lagnado et al. (2006) used 50-trial windows in their analysis, while Meeter et al. (2006) used windows of 24 trials. To equate the two, we ran both with 30-trial windows (i.e., $d = 30$), although for both, d did not matter much for the quality of the predictions (see Fig. 5). This means that only the trials after the 30th were analyzed, as the first 30 trials were necessary to generate predictions for trial 31.

Fig. 5 shows that the strategy analysis made fewer prediction errors than the rolling regression analysis. The comparison is unfair, however, as strategies often predict random behavior on certain patterns, and thus in fact do not make a prediction for trials in which one of these patterns was presented. By contrast, even when the rolling regression analysis gives a likelihood of responding "rain" of 0.51 (i.e., very close to chance level), this was counted as a prediction of rain. Indeed, the likelihood that a prediction of the rolling regression analysis was correct was dependent on the distance from 0.5—e.g., a trial in which the rolling regression analysis predicted a "rain" response with a likelihood of 0.8 was more likely to indeed have had a rain response than a trial on which the predicted likelihood was 0.6 (see Fig. 6). To correct for this, we added one half of the trials in which no prediction was made by the strategy analysis to the error trials. With this correction, the two forms of analysis seem largely equivalent in their ability to predict responses.

Given that the two forms of analysis are based on very different assumptions, it is surprising that they are so similar in their ability to predict responses. One possibility is that some participants engage in rule-based learning and others in incremental learning, as has been shown in deterministic categorization tasks (Raijmakers et al., 2001). Some participants would then be well characterized by strategy, others by rolling regression analysis. This turns out not to be the case: overwhelmingly, the same participants who are characterized correctly (or incorrectly) by one analysis tend to be characterized correctly (or incorrectly) by the other analysis (see Fig. 7). The correlation between the number of responses predicted well by the one or the other was 0.95. It was also not the case that rolling regression was superior in predicting early trials and strategy analysis in later trials. Both were better at predicting responses in the second half of the experiment (trials 101–200) than in the first half (trials 31–100), $F(1,29) = 8.33$, $p < 0.001$, but no interaction was
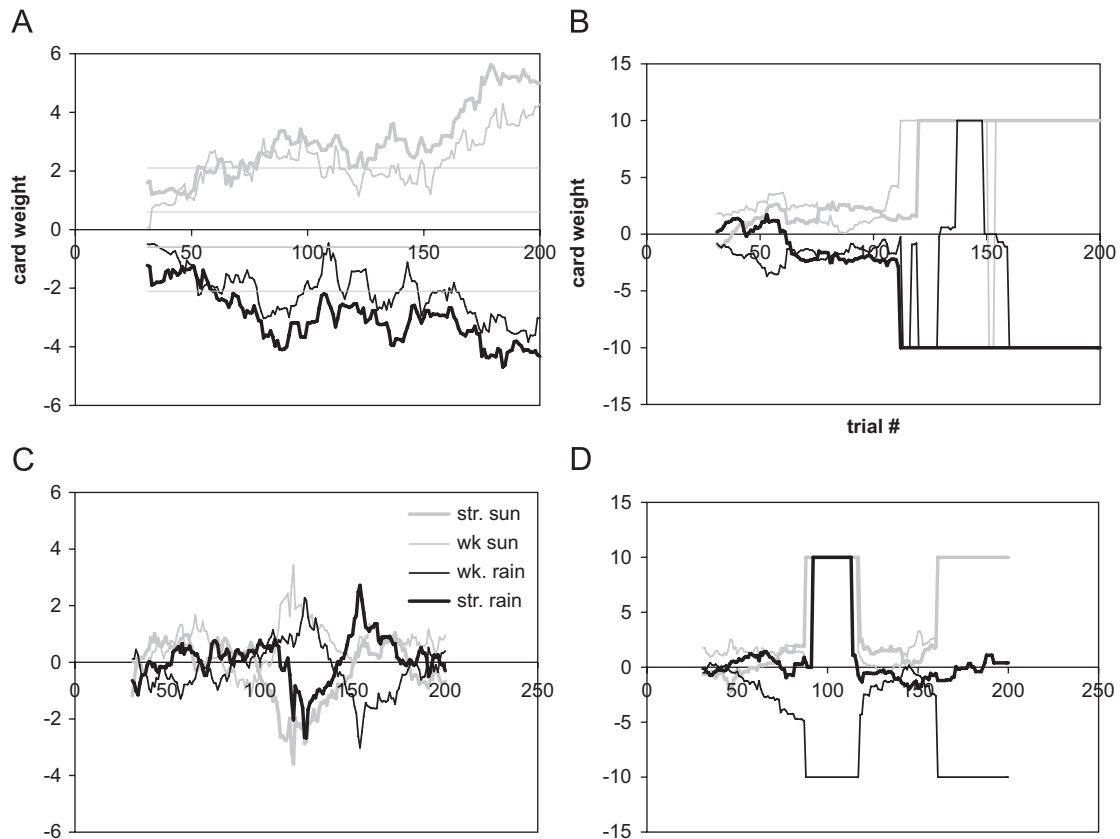
Fig. 4. Outcomes of the rolling regression analysis for the whole group, and for the three participants analyzed in Fig. 3B. (A) Average regression weights of the four cards for the 30 participants after capping weights at 10. Gray lines give the objective card weights of the cards, as would be predicted by the true predictive power of the cards (from top to bottom: the objective weight of the strong sun, weak sun, weak rain and strong rain card). Other panels give the regression weights computed from participant 1 (B), participant 2 (C) and participant 3 (D).
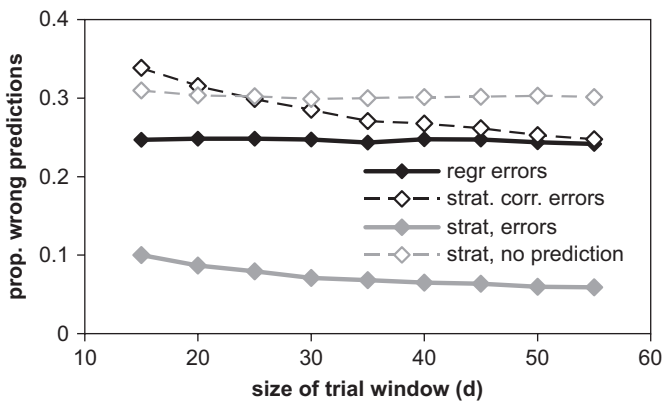

Fig. 5. Proportion of trials predicted well by strategy analysis and by rolling regression analysis, as a function of the number trials used to generate the prediction with (referred to as *d* in the text). Three lines are given for the strategy analysis: one for the proportion of trials where the prediction was wrong, one for the proportion of trials for which no prediction was made. The dashed black line combines trials with wrong predictions and half of the trials in which no prediction was made.

Fig. 6. Proportion of trials on which participants responded "sun", as a function of the likelihood of such an answer predicted by rolling regression and strategy analyses. For the rolling regression analysis, predictions were grouped into bins of 0.1, labeled by their midpoint. Strategy analysis generates only three predictions: "sun", "rain" or "guess". These are marked by three boxes.

found between experiment half and type of analysis, $F < 1$ (there was also no main effect of type of analysis, $F(1, 29) = 1.93$, $p > 0.1$).

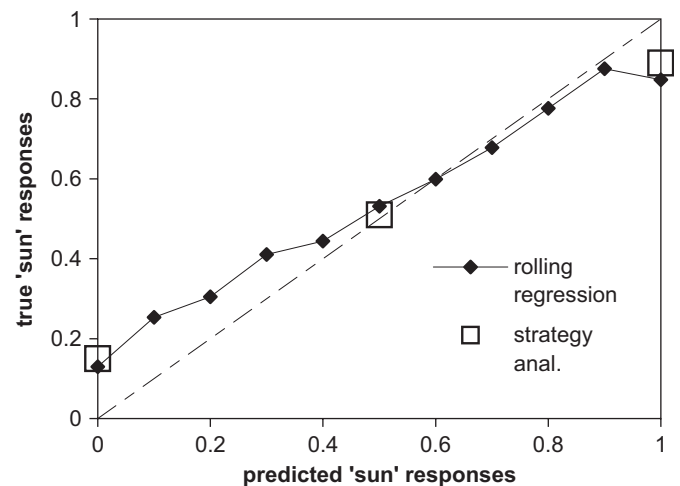A reason for the similar performance of the two analyses is that their predictions are highly correlated. Table 1

shows trials of all participants separated out by prediction of the rolling regression analysis and of the strategy analysis. The table shows both the number of trials in each
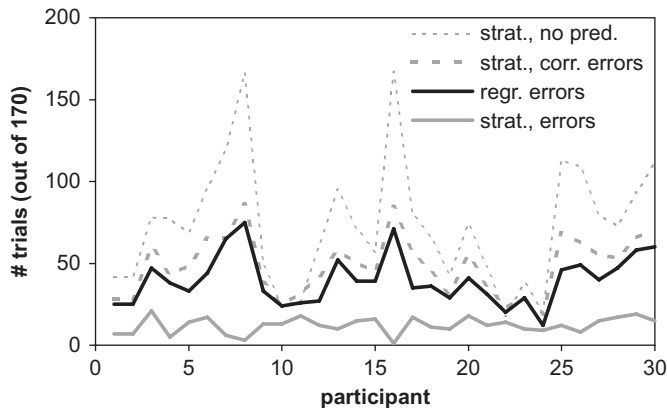
Fig. 7. For each of the 30 participants, the proportion of trials predicted well by strategy analysis and by rolling regression analysis. Three lines are given for the strategy analysis: one for the proportion of trials where the prediction was wrong, one for the proportion of trials for which no prediction was made. The last line combines trials with wrong predictions and half of the trials in which no prediction was made.

Table 1
The proportion of trials on which a "sun" response was given, as a function of the predictions made by strategy analysis (rows) and rolling regression analysis (columns)

	Rolling regression prediction		
	Rain response	Sun response	Average
Strategy prediction			
Rain response	0.14 (25%)	0.33 (2%)	0.15
Sun response	0.65 (1%)	0.90 (26%)	0.89
No prediction	0.38 (22%)	0.62 (24%)	0.51
Average	0.26	0.75	

Also given, in the parentheses, is the proportion of trials in each category.

cell (as a percentage of all trials), and also the proportion of trials in which the respondent answered "sun". From the table, it is evident that on most trials the predictions of the two forms of analysis are the same. For those trials in which the two analyses make opposite predictions, responses are most often in line with the prediction made by the strategy analysis (i.e., the proportion of "sun" responses is high for trials in which strategy analysis predicts sun but rolling regression analysis rain, and vice versa). On the other hand, on the trials on which strategy analysis does not make a prediction, the prediction of the rolling regression analysis is clearly above chance. These findings suggest that strategy analysis is better at capturing stable patterns in performance, but that some learning is overlooked which is captured by the rolling regression analysis. These responses are not characterized very well by rolling regression analysis either, however, as the odds of a "sun" response is 1.63 in the trials in which rolling regression analysis predicts "sun" and strategy analysis makes no prediction.

4.3. Jerkiness of learning

A difference between the assumed learning profiles is that strategy analysis assumes a "jerky" learning trajectory, while rolling regression analysis is based on the idea of smooth, incremental learning. As shown in Fig. 4B and D, in some participants cards weights change dramatically from one trial to the next, more suggestive of a "jerky" learning trajectory than of incremental learning. Such jumps in card weights seem indicative of strategy switches, which are indeed diagnosed by the strategy analysis at about the same trial. We investigated whether such jumps in cards weights occurred more often than could be expected if learning were gradual.

To provide a fair baseline, we generated a simulated weight change for each transition from trial n to trial $n+1$, assuming gradual weight change. Again, 30-trial windows were used. We took the fitted weights at the start and the end of the windows used to fit trial n and $n+1$, and then assumed a linear change of weights over the course of the 30 trials (i.e., if the weight for a cue increased from 0.2 on the first trial of the window to 0.5 on the 30th trial, the weight for this cue on trial 5 in the window was assumed to be 0.25). For each trial, we then computed from the weights a likelihood of responding "sun" or "rain" given the pattern that, in that trial, was being presented to the observer. Monte Carlo simulation was then used to create a concrete set of responses from the likelihoods. These were fitted using rolling regression, resulting in weights for trial n, and weights for trial $n+1$. The difference between these two sets of weights was taken as the baseline for the weight changes observed in the data. Both the real and the baseline weights were first capped at 10 or –10 before weight changes were computed.

Fig. 8 shows the distribution of weight changes over the trials 51–150 in the experimental data, and in the simulated weight change baseline. Larger weight changes (i.e., larger than 0.10) were more frequent in the data than in the baseline, $t(29) > 8$, $p < 0.001$ for all four cues. In a further difference, changes of the weights of individual cards are correlated in the data (r ranging from 0.07 for the strong and weak "sun" cue, to -0.486 for the two strong cues, all $p < 0.001$); this is not the case in the baseline changes (r ranging from 0 to 0.04, all insignificant by the criterion of 0.009 set by the Bonferroni correction). Correlated weight change would be expected if sudden strategy shifts underlie the weight changes, as in such shifts weights of all cues would shift at once.

4.4. Conclusion

Strategy analysis and rolling regression analysis are more or less equivalent in the ability to predict responses in healthy subjects. Neither is particularly good as a predictive tool: for rolling regression analysis, approximately 30% of responses were counter to the prediction. Strategy analysis led to fewer faulty predictions, but in
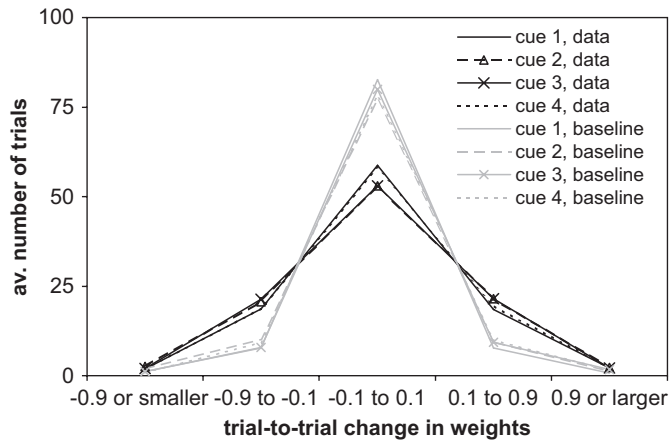
Fig. 8. Distribution of the change in regression weights from trial to trial, averaged over 30 participants. Black lines ("data") show the change in weights fitted to the responses of the participants, for trials 51–150. Gray lines ("baseline") show what the distribution would be if weights used by the participant to generate responses changed incrementally over the course of the experiment. See main text for how the baseline weight change distribution was derived.

many trials no prediction was made. Correcting for the trials with no prediction led to approximately the same error rate as rolling regression analysis. This was the case at a collective level, but also at an individual level. Some participants respond in a predictable fashion to the patterns, others do not, and this is independent of the analysis used.

There are several ways in which this data pattern can be understood. First, it is possible that the rule-based learning analyzed with strategy analysis and the incremental learning analyzed with regression are present in all participants, and are highly correlated. In this case, a combination of both types of analysis would be best at characterizing learning. Second, it is possible that one form of analysis mimics the other. For example, it could be that the regression analysis captures true learning, and that the strategy analysis is able to capture variance in the learning, because certain sets of weights resemble strategies.

A suggestion of regression analysis mimicking strategy shifts can be found in sudden weight changes. At trials in which the strategy analysis identifies a strategy switch, large changes of cue weight are found by regression analysis. Such large changes occur with higher frequency than can be explained by incremental change in underlying weights used by participants. There are several caveats in interpreting this result, however. First, the weights and strategies were derived from windows of trials, which by necessity entails a smoothing of the learning trajectory. If, for example, the strategy of a participant changes at trial 63, this will not immediately become apparent in the weights for trial 64, as these are based on trials 35–64. On most of these trials, the "old" strategy was still used. A second problem is that a less smooth progression of weights than assumed here might produce more large

weight changes, in line with what is found in the data. Future studies will need to determine to what extent the "jerkiness" of the learning trajectory is real, and to what extent it is a methodological artifact.

5. Patient populations and probabilistic categorization

We now turn to evidence from patient populations. To reiterate, if probabilistic categorization would rely on procedural learning, the task could be assumed to rely on the basal ganglia. Rule-based learning would imply a reliance on both the basal ganglia and the frontal cortex, while instance memorization would imply a reliance on the medial temporal lobe. We will discuss evidence with regard to each region in turn.

5.1. Basal ganglia

Several papers have investigated probabilistic categorization in patients with basal ganglia abnormalities. These studies have predominantly included participants with Parkinson's disease. In Parkinson's disease, dopaminergic projections to the basal ganglia are affected, leading to abnormal basal ganglia functioning. As is reviewed in more detail elsewhere in this issue, patients with Parkinson's disease are consistently impaired on probabilistic categorization tasks. Although participants with Parkinson's disease do achieve above-chance performance, their learning is significantly slower than that of matched controls. (Shohamy et al., 2004; Perretta et al., 2005). In a somewhat different study, degeneration of brain tissue in patients with Alzheimer's disease was measured through magnetic resonance spectroscopic imaging (MRSI). Basal ganglia degeneration correlated with poor performance in the Weather Prediction task (Colla et al., 2003).

Corroborating evidence for a role of the basal ganglia in category learning comes from several studies that assessed brain function of healthy volunteers via functional magnetic resonance imaging (fMRI), while the participants performed the weather prediction task. All found increased activity in the striatum of the basal ganglia after the first few trials (Poldrack et al., 2001; Aron et al., 2004; Foerde et al., 2006).

5.2. Prefrontal cortex

Whereas results on basal ganglia involvement are quite clear-cut, this is not the case for the prefrontal cortex. One study investigated a population with varying prefrontal lesions, and did not find any effects of these lesions on probabilistic categorization (Perretta et al., 2005), although this may have been due to variability in the location of lesions. By contrast, fMRI studies do suggest involvement of prefrontal cortical areas in probabilistic categorization (Aron et al., 2004; Flanery, 2005). In these studies, probabilistic categorization was compared either

with a memorization task (Aron et al., 2004) or with a prototype distortion categorization task (Flanery, 2005). Furthermore, a transcranial magnetic stimulation (TMS) study found that stimulating the prefrontal cortex (in a regime thought to enhance excitatory transmission) was beneficial for probabilistic categorization performance (Kineses et al., 2004). Whereas lesion evidence thus suggests only a limited role of the prefrontal cortex in probabilistic categorization, imaging and TMS studies suggest the opposite.

5.3. Medial temporal lobe lesions

Whether amnesic patients with hippocampal damage are impaired at probabilistic category learning has been debated in the literature. The first report with amnesic patients of mixed etiology (including hippocampal and diencephalic patients) found no learning impairment relative to healthy controls early in learning, although an impairment did emerge with extended training (Knowlton et al., 1994). Moreover, patients with Alzheimer's disease were found to be unimpaired in weather prediction (Eldridge et al., 2002). A later report considering only amnesic patients with bilateral hippocampal damage due to hypoxic brain injury found that amnesic patients were impaired both early and late in learning (Hopkins et al., 2004). This latter paper conducted strategy analyses on the amnesic patients and controls, and suggested that the amnesic patients did not use complex strategies as often as control participants did.

The strategy analyses reported by Hopkins et al. (2004) could not determine, however, whether the amnesic patients failed to acquire *any* strategy, whether they acquired a simple strategy but then did not switch to a complex one later, or whether they could acquire a complex strategy but abandoned it more often than control participants. Meeter et al. (2006) reanalyzed the data from Hopkins et al. (2004) to more precisely determine the cause of the learning decrements in the amnesic patients. Meeter et al. (2006) found that at the start of learning, there was little that differentiated the performance of amnesic patients and control participants. In both groups, some participants adopted simple strategies whereas others did not. Nevertheless, differences in strategy use between amnesic patients and control participants appeared quite early in learning (i.e., around trial 40). Normal controls gradually moved to more complex strategies. Amnesic patients, on the other hand, fell back to having no recognizable strategy as often as they switched to a different strategy. These findings suggest that the amnesic patients are unable to keep track of attempted strategies and of the feedback received over the course of the experiment. Such an inability to remember the strategies or feedback would fit with the general pattern of abnormally rapid forgetting in amnesic patients.

In apparent contrast with the clinical results, evidence from functional imaging (fMRI) is suggestive of an MTL role early in learning. In one study, MTL activity was only observed in the first 15 trials of the weather task. After these trials, activity in the MTL was even lower in probabilistic categorization than in the control condition. (Poldrack et al., 2001; Aron et al., 2004). Foerde et al. (2006) provided evidence that this activity is not artifactual. They subjected participants to a distracting secondary task while performing weather prediction. In a control condition in which participants concentrated on weather prediction, activity in the MTL was predictive of performance in a probe phase in which no feedback was given. When participants were distracted, performance was as good as in the control condition, but now activity in the basal ganglia, and not in the MTL, predicted performance. Moreover, explicit knowledge of the task, as measured by a questionnaire, was lower after distraction than in the control condition. This suggests that during normal task performance, the MTL is involved in performance, even though it is not necessary as shown by normal performance in the distraction condition.

6. Discussion

A typology used in deterministic categorization identifies three ways in which to approach a categorization task, with the categorization structure determining which of the three was chosen by most categorizers (Ashby and Ell, 2001). Rule-based learning, reliant on the frontal lobes and the basal ganglia, consists of trying to find rules to base categorization on. Incremental learning, reliant on the basal ganglia, consists of finding category boundaries in the stimulus space slowly over trials. The third approach consists of simply memorizing all exemplars in the task, together with their category assignment. This kind of learning relies on the temporal lobes. Here, we investigated which of these three offers the best description of learning in probabilistic category learning. Two kinds of evidence were considered: performance of normal categorizers, and data from clinical populations.

Two characteristics of learning could help determine how normal categorizers approach probabilistic categorization. First, different types of analysis are suited to detect different kinds of learning: strategy analysis for rule-based learning, rolling regression analysis for incremental learning. If one type of analysis were better at characterizing performance, this would suggest that its associated kind of learning is more prominent than the other kind. In fact, both types of analysis proved equally good at predicting performance. Second, rule-based learning predicts "jerky" learning, with sudden jumps in performance whenever the categorizer adopts a new rule or strategy. By contrast, incremental learning assumes a smooth progression of learning over the course of the experiment. Here, evidence was found for a jerky progression of learning, but more studies are clearly needed to draw firm conclusions.

We then turned to evidence from cognitive neuroscience and neuropsychology. Ample evidence was found for an involvement of the basal ganglia in probabilistic categorization. Evidence for a prefrontal involvement was mixed. Although patient data was not suggestive of a strong prefrontal role, data from imaging and TMS studies did suggest that the prefrontal cortex is important in probabilistic categorization. Rule-based learning would suggest a reliance of probabilistic categorization on the prefrontal cortex, while incremental learning would not. Neither option is thus falsified by evidence from cognitive neuroscience and neuropsychology.

Neither option would prima facie predict on involvement of the medial temporal lobe in probabilistic categorization. Yet, evidence from patient and imaging studies clearly suggest such a role. This could point to a role of memorizing instances in probabilistic categorization. In this case, however, only deficits late in learning would be expected in patients with MTL damage (Ashby and Ell, 2001), whereas deficits in patients are apparent relatively early in the task (Hopkins et al., 2004), as is MTL activity as evidenced by fMRI (Poldrack et al., 2001). Moreover, if participants memorize patterns and their category assignment, the response to pattern i presented on trial t should be the same as previous responses to that same pattern. Meeter et al. (2006) found, however, that the response on to a pattern i was predicted relatively badly by the response given on the previous trial with pattern i. This suggests that memorizing category assignments of individual patterns did not play a very large role in the performance of the participants.

All data together give the impression that a contrasting rule-based learning, incremental learning and memorizing pattern–category pairings are not sufficient to understand probabilistic categorization. Two aspects of the data suggest that a new synthesis, involving all three, might give a better account of learning in such tasks. First, it was found that individual participants are characterized well to the same extent by strategy analysis and rolling regression analysis. Whereas strategy analysis gave a better account of robust patterns in performance, the rolling regression analysis identified patterns in behavior in trials that the strategy analysis does not give a prediction for. Participants may thus engage in both types of learning, with strategy analysis picking up rule-based components and rolling regression incremental learning that occurs at the same time and perhaps underlies the formation of rules. Third, fMRI data shows high MTL involvement and low basal ganglia activity early in learning, but high basal ganglia and low MTL activity late in learning (Poldrack et al., 2001). This can be taken to suggest that the MTL helps set up representations of the stimulus set early in learning. These are then used by other brain areas (such as basal ganglia or the prefrontal cortex) to assign patterns to categories later in learning. Such a role—setting up the right representations for other brain areas to use—has been proposed for the MTL in a computational model of classical condition-

ing (Gluck and Myers, 1993, 2001). Similar computational work could shed a new light on how brain regions together underlie probabilistic categorization.

References

Aron, A.R., Shohamy, D., Clark, J., Myers, C.E., Gluck, M.A., Poldrack, R.A., 2004. Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. Journal of Neurophysiology 92, 1144–1152.

Ashby, F.G., Ell, S.W., 2001. The neurobiology of human category learning. Trends in Cognitive Sciences 5, 204–210.

Cohen, N., Squire, L., 1980. Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science 210, 207–210.

Colla, M., Ende, G., Bohrer, M., Deuschle, M., Kronenberg, G., Henn, F., Heuser, I., 2003. MR spectroscopy in Alzheimer's disease: gender differences in probabilistic learning capacity. Neurobiology of Aging 24, 545–552.

Eldridge, L.L., Masterman, D., Knowlton, B.J., 2002. Intact implicit habit learning in Alzheimer's disease. Behavioral Neuroscience 116, 722–726.

Flanery, M.A., 2005. The neural correlates of explicit categorization. Department of Psychology, Vanderbild University, Nashville, TN.

Foerde, K., Knowlton, B.J., Poldrack, R.A., 2006. Modulation of competing memory systems by distraction. Proceedings of the National Academy of Sciences USA 103, 11778–11783.

Gluck, M.A., Myers, C.E., 1993. Hippocampal mediation of stimulus representation: a computational theory. Hippocampus 3, 491–516.

Gluck, M.A., Myers, C.E., 2001. Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus in Learning and Memory. MIT Press, Cambridge, MA.

Gluck, M.A., Shohamy, D., Myers, C.E., 2002. How do people solve the "weather prediction" task? Individual variability in strategies for probabilistic category learning. Learning & Memory 9, 408–418.

Graf, P., Schacter, D.L., 1985. Implicit and explicit memory for new associations in normal and amnesic subjects. Journal of Experimental Psychology: Learning, Memory and Cognition 11, 501–518.

Hopkins, R.O., Myers, C.E., Shohamy, D., Grossman, S., Gluck, M.A., 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.

Kineses, T.Z., Antal, A., Nitsche, M.A., Bártfai, O., Paulus, W., 2004. Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the human. Neuropsychologia 42, 113–117.

Knowlton, B.J., Squire, L.R., Gluck, M.A., 1994. Probabilistic classification learning in amnesia. Learning & Memory 1, 1–15.

Knowlton, B.J., Squire, L., Paulsen, J., Swerdlow, N., Swenson, M., Butters, N., 1996. Dissociations within nondeclarative memory in Huntington's disease. Neuropsychology 10, 538–548.

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.

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, 230–239.

Moody, T.D., Bookheimer, S.Y., Vanek, Z., Knowlton, B.J., 2004. An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. Behavioral Neuroscience 118, 438–442.

Perretta, J.G., Pari, G., Beninger, R.J., 2005. Effects of Parkinson disease on two putative nondeclarative learning tasks: probabilistic classification and gambling. Cognitive and Behavioral Neurology 18, 185–192.

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

Raijmakers, M.E.J., Dolan, C.V., Molenaar, P.C.M., 2001. Finite mixture distribution models of simple discrimination learning. Memory & Cognition 29, 659–677.

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.

Squire, L.R., 2004. Memory systems of the brain: a brief history and current perspective. Neurobiology of Learning and Memory 82, 171–177.

Sun, R., Slusarz, P., Terry, C., 2005. The interaction of the explicit and the implicit in skill learning: a dual process approach. Psychological Review 112, 159–192.