Decision Making in Uncertain Situations

Elaine Duffin

MSc Artificial Intelligence

Session 2010/2011

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

(Signature of student)__________________________

Abstract

Human ability to respond to data which has both random fluctuations and systematic changes was studied by Bland and Schaefer at the University of Leeds [1]. This work examines the data from that study. It was found that there was more than one style of learning behaviour shown by the human participants.

Behrens *et al* [2] considered human learning in a similar study to that of Bland and Schaefer. The learning of a Bayesian network was compared to that of the human participants. Bayesian networks are a machine learning technique which can allow optimal responses to noisy data. This study builds and experiments with computational models based on the work of Behrens *et al*. It was found that the model used by Behrens *et al* was more complex than required for the learning task. Computational models are compared to the human participants of the study of Bland and Schaefer.

Acknowledgements

I have a great deal for which to thank my supervisor, Dr Marc de Kamps. Marc has given positive comments when I have felt that my work was not going well. He has provided endless encouragement during long meetings.

I thank Dr Alexandre Schaefer of Durham University for his time and discussions on the psychological side of the study.

I also thank my fellow student Aswathy Kishore for her helpful comments following a careful proofreading of this report.

Contents

Chapter 1

Introduction

1.1 Motivation

People often have to learn in situations in which they are not directly told what they need to learn, they have to learn from experience. Experiences may conform to a basic underlying rule which can be learnt, but there may be departures from the rule. Psychologists study learning using experiments which control the experiences people receive. Some such studies are described in Chapter 2.

Experimental work into human learning was carried out by Bland and Schaefer [1, 3] in the Institute of Psychological Sciences at the University of Leeds. In this work, participants had to respond to the colour of a triangle presented on a screen by pressing one of two buttons. The participants were each presented with 960 successive stimuli and given immediate feedback as to whether each response was correct. The feedback reflected an underlying rule but occasionally went against that rule. Different experimental conditions were created by altering the proportion of feedback which opposed the current underlying rule and by changing the underlying rule. Details of the experimental paradigm can be read in Chapter 2.

The motivation for the current work is to test the hypothesis that humans use Bayesian reasoning when learning. Bayesian reasoning allows the updating of existing beliefs with new experiences and can produce optimal performance. The experimental data from the study by Bland and Schaefer were made available for the present work and are used in comparison with computational Bayesian models.

The experimental data of Bland and Schaefer are analysed to look for features in the behaviour of the participants in different experimental conditions. This analysis shows that people do not behave in an optimal way, as described in Chapter 3. The findings will be compared to those of other psychological studies.

Previous research has sought to understand how learning takes place in similar experiments to that of Bland and Schaefer. Behrens *et al* [2] created a Bayesian learning model to predict outcomes in a learning experiment. They found that the behaviour of the human participants more closely matched that of the Bayesian learner than an alternative non-Bayesian learning model. They took this as evidence for Bayesian learning in humans.

The current study aims to implement a Bayesian learner based on that of Behrens *et al.* This computational model will be compared with the responses of the human participants provided by Bland and Schaefer. The model of Behrens *et al* is complex, having several variables. The current work aimed to discover whether this complex model performs better in the learning task than simpler models with fewer variables. A hierarchy of models is defined with increasing levels of complexity starting from a simple application of Bayes' theorem and working up to the model used by Behrens *et al.* A description of the models used can be found in Chapters 4 and 5.

Although the human behaviour had been found not to be optimal, it was possible that there would be features of the human behaviour which might indicate Bayesian learning. In particular, I wished to compare the behaviour of humans and the computational model in the period immediately following a change in the underlying rule. Bland and Schaefer reported on the behaviour of human participants in response to negative feedback and found significant differences according to experimental conditions. I wanted to see if the same differences occurred when using the model used by Behrens *et al.* Although the performance of the human subjects is clearly not optimal, I wanted to investigate whether this feature was shared by both the human and Bayesian learners. The comparisons between the human and Bayesian models are described in Chapter 6.

A discussion on the behaviour shown by humans and the performance of the computational models will appear in Chapter 7.

1.2 Methodology

I chose to manage my project using an iterative approach. This meant that I chose initially to plan the work for the first part of the project, up to the requirement to submit an interim report. This was to include background reading, modelling, evaluation and write-up. The initial plan can be seen in Appendix D. I decided not to create a detailed plan covering the whole project as it was difficult to estimate the time required for research activities. The original project plan assumed an iterative approach with steps of background reading, programming and evaluation repeated multiple times throughout the project. Planning would then be carried out for the later phase.

I initially believed that the experiment carried out in Leeds by Bland and Schaefer was very similar to that done by Behrens *et al* and that I would be merely applying their computational methods to the new data. Behrens *et al* claim that Bayesian learning is optimal in uncertain situations and that there is evidence of Bayesian learning in the experimental data they obtained.

My analysis of the data of Bland and Schaefer suggests that there are at least two types of behaviour shown in the study. One behaviour, probability matching, is definitely not optimal. This behaviour has been observed in many psychological studies. Only a few participants showed behaviour which maximised their rewards. There were not enough participants to draw any clear conclusions about Bayesian behaviour.

My findings from the experimental data have led to a change in focus for the computational modelling part of the project. I had concerns about the complexity of the number of conditions in the study by Bland and Schaefer and so decided that for comparing computational models it would be better to generate new test data more like that used by Behrens *et al*. This also allows me to have much more data available on which to test computational models compared to the data for a psychological study.

In addition to the creation of data on which to test computational models, I decided to create multiple models with differing levels of complexity. This increased the time required for modelling and evaluation. A provisional plan was made and submitted with the interim report, taking account of these changes. The plans can be seen in Appendix D.

After submission of the interim report, a meeting was held between Dr Alexandre Schaefer, Dr Marc de Kamps and myself at Durham University. Bland and Schaefer had made revisions to their report which had been submitted for publication. These revisions included additional analysis of the behavioural responses of the participants and findings of significant differences in behaviour in different experimental conditions. Although I was already aware that humans were not showing full Bayesian decision making, I wanted to see whether any of these new findings applied to Bayesian learners. This allowed me to modify my plan to cover the remainder of the project. The new plan can also be seen in Appendix D.

Chapter 2

Background

2.1 Human learning in an uncertain environment

Many studies have involved humans learning a rule from experience. In one type of study, participants have to predict which of multiple outcomes will occur next over a sequence of trials. They are not told what rules apply, but are expected to update their beliefs by taking account of the results of previous trials. There is often a random element to the sequence of outcomes, an uncertain environment.

In Siegel and Goldstein, 1959 [4], participants had to predict which of two lights, left or right, would next be illuminated in a sequence of 100 trials. The order of illumination of the lights was determined randomly, but with one light occurring three times more frequently than the other. The randomisation was determined using random number tables but with constraints that there were no more than six consecutive presentations of the more frequent event, and that in every block of 20 trials, the lights were in the exact 3:1 ratio. After making a prediction, the participant saw a light come on, thus gaining immediate feedback as to whether their prediction was correct. The participants were asked to try to make as many correct guesses as possible.

Siegel and Goldstein discuss two models which predict the behaviour of the participants.

- The ratios of predictions of left or right match the ratios of illumination of the lights. This is a commonly observed phenomenon and is often referred to as probability matching ([5, 6]).
- The participants learn which light is illuminated more frequently and after learning the rule, they predict the more frequent event on all trials. This behaviour is referred to as maximising [5].

As the sequence of illumination of the lights is random, maximising behaviour is optimal and results in more correct responses than does probability matching. Given a random sequence, participants would not be able to predict the occurrence of the less likely outcome and would benefit from sticking with the known rule to guarantee success on at least those trials.

Decision making theory, (for a discussion see [5, 7]) suggests that people will behave in a way which maximises a value which is referred to as utility. This value is subjective, it depends on the individual involved. Siegel and Goldstein suggest that probability matching may be a form of maximising if the participants feel more satisfaction when they correctly predict the less frequent event or that it might relieve boredom of always predicting the same light. They hypothesised that if there was a monetary reward for correct responses, then this incentive might override satisfaction from other factors.

Siegel and Goldstein carried out a variation of the experiment, using three groups of participants. One group, referred to as the reward group, received a payment of five cents for every correctly predicted light. The second group, the risk group, also received the payment for correct predictions but in addition they lost five cents for every incorrect prediction. The third group, the no payoff group, undertook the basic procedure with no monetary gains or losses arising from their predictions.

Siegel and Goldstein looked at the number of times that the more frequent light was predicted during the last 20 trials. The results were that the more frequent light was predicted more in the risk group, followed by the reward group and then by the no payoff group. They carried out some additional trials with a small number of participants giving results which suggested that learning was faster in the risk group.

Following many studies which reported probability matching behaviour, Shanks *et al* [6] raised the following objections to the view that people behave irrationally by probability matching.

- The sequence of events used is not always completely random. For example, the constraints used by Siegel and Goldsmith described above, lead to a non-random sequence. If there is a pattern to the presentation, then the participants might be able to work out when to choose the less frequent event.
- Many studies did not provide monetary rewards for correct predictions or the monetary rewards were of a very low value. Shanks *et al* report that some studies suggest that maximising behaviour is observed more when monetary rewards are given.
- Participants may be trying to find patterns in the sequence of outcomes.
- Most studies have reported group rather than individual behaviour. This can obscure individual differences between participants so that individuals who maximise may not be identified.

Shanks *et al* rejected the claim that probability matching could be explained as maximising if relieving boredom was considered to be a reward. They believed that if this were the case then offering a large enough monetary reward would outweigh any other factors and allow maximising to be shown.

Some of the above issues were addressed in a series of experiments carried out by Shanks *et al.* They gave monetary rewards for correct predictions and told the participants that there was no pattern but that they could do well if they thought about their strategy. They gave feedback on how well the participants were doing in comparison to an optimal strategy, without instructing the participants about how an optimal strategy would make decisions. Their first experiment consisted of 300 trials, but they found that, for some of the participants, the behaviour had not reached an asymptote. They repeated the experiment using 1800 trials. They defined maximising as having been shown if a participant chose the more frequent option every time for a full block of 50 trials. Using this definition, they found that 8 out of 12 participants were maximising and the other 4 were probability matching.

Gaissmaier and Schooler [8] suggested that probability matching arises from the participants searching for patterns. They set up experiments to investigate probability matching in a prediction task of 576 trials which were split into two blocks. In one block of 288 trials, the presentation was random and the more frequent event occurred 67% of the time. The other block of 288 trials consisted of a repeating pattern of length 12, the more frequent event again occurring in 67% of trials.

From the behaviour in the random block, they split the participants into matchers and maximisers. They then looked at the behaviour in the pattern block. The participants identified as matchers in the random block improved more over time than the maximisers when presented with events following a pattern. In this case, it was an advantage to search for patterns.

Gaissmaier and Schooler concluded that probability matching was a result of pattern search for some participants but of a win-stay, lose-shift strategy for others. A win-stay, lose-shift strategy involves using the same rule to make a decision as the previous trial if the previous decision was rewarded, otherwise changing decision.

Koehler and James [9] carried out a study to examine the extent to which probability matching occurred even when pattern information could not be used. They held an alternative view to that of Gaissmaier and Schooler, that probability matching was a mistake in reasoning rather than a search for patterns. They believed that people keep track of probabilities of different outcomes but then use those probabilities to choose their predictions. People generally find it easier to state how many of an outcome should be expected than to work out what would be the best way to guess an outcome. They claim that people either do not consider alternative strategies or do not choose a maximising strategy.

Koehler and James set up a study in which red or green marbles were pulled from a bag. The study was split into a learning phase and a test phase, each of which could be carried out in a serial or aggregate manner. In serial learning, the participants saw marbles being drawn randomly one by one from a bag. In aggregate learning, the participants were told how many red and how many green marbles had been drawn without seeing the individual draws. In the test phase, the participants were told that they would earn money for correctly guessed colours. The serial test condition required the participants to predict a colour one marble at a time. In the aggregate test condition, the participants simply had to give a predicted total number of each colour. Many more of the participants used a probability matching strategy (46 participants) than used maximising (14 participants). The rest of the 102 participants used a strategy which was rated to be between probability matching and maximising but closer to probability matching. Koehler and James found no evidence that matching was more likely in the serial conditions in which pattern information was available.

After the marble guessing, the participants were given a questionnaire in which they had to evaluate and rank different possible strategies for the task. In the questionnaire, 50 participants ranked maximising higher than probability matching, many more than had used a maximising strategy in the marbles task. Even when answering the questionnaire, slightly more than half (52) of the participants still ranked probability matching higher than maximising. Koehler and James believed that maximising was not a strategy that is considered by most people.

Behrens *et al* [2] asked participants to choose between two colours. The probability of each colour being correct varied during the experiment. During what was referred to as a stable period, one colour appeared 75% of the time. A volatile period consisted of blocks of 30 or 40 trials with the more frequent colour appearing 80% of the time but with the rule reversed in each block. The value of the reward for a correct decision varied between trials and was displayed to the participants. Behrens *et al* compared the behaviour of the participants to a computational Bayesian learner and claimed that the human behaviour was close to being optimal. Behrens *et al* noted that the participants often chose the less likely colour, but attributed this behaviour to the reward value being larger for the less likely colour on those trials.

In the experiment by Bland and Schaefer [1], participants saw a red or blue triangle on a screen on each trial and had to respond by pressing one of two buttons. After responding, the participants were presented with a message informing them whether or not they were correct. The correct answer was based on an underlying rule, for example when red press 1, when blue press 2. On a random selection of trials, the opposite response would be classed as correct, the proportion of such trials was determined by an error rate. The error rate had two different levels, high and low, but remained constant during a block of 120 trials. The participants started with 1000 points and gained 10 points for every correct response and lost 10 points for every incorrect response or when they failed to respond within 1000 ms. The participants were not given any information about the underlying rules but were asked to try to gain as many points as possible which would be converted into money at the end of the experiment. As with Behrens *et al*, blocks were split into volatile and non-volatile blocks.

Bland and Schaefer took EEG data during the study, in addition to recording the responses of the participants. They expected to see evidence in the EEG signals of different cognitive processes to track changing rules in low error compared to high error conditions. They concluded that this hypothesis was supported by the EEG data.

2.2 Probabilistic Reasoning

2.2.1 Introduction

Uncertainty can be considered to cover two different situations, firstly, those in which not all information is available and secondly, where a non-deterministic process is involved, as described by Russell and Norvig [10]. In the psychological experiments considered in Section 2.1, both of these apply; the participants cannot directly observe the underlying rule and the outcome is determined in a stochastic way from the rule.

Cox [11] demonstrated that the rules of probability could be applied to beliefs as well as to countable frequencies. Bayesian reasoning provides an optimal way to update beliefs based on evidence. The beliefs can then be used to make predictions for future events. Using information from observations to reason about other beliefs is known as probabilistic inference.

2.2.2 Terms and Notation used

In probability, a random variable is one which can exist in one of a finite number of states which are mutually exclusive. Associated with a random variable is a probability distribution which is an ordered list of values corresponding to each state and representing the probability that the variable is in that state. Each of these values, being a probability, must lie between 0 and 1. The sum of these values must be 1 as the variable has to be in one of those states.

A simple example of a random variable could be one which represents the probability that the grass is wet. This random variable could take two states, true and false. If the random variable is represented by G, then the probability distribution over G is a vector of values $(P(G=t), P(G=f))$. This vector consists of the probability that the grass is wet and the probability that the grass is not wet respectively. This probability distribution will be written as $\mathbf{P}(G)$. This use of a bold font to denote a vector or table of probabilities is taken from the notation used by Russell and Norvig.

A random variable has an expected value, or mean, which is an average of the possible values of the random variable weighted by their probabilities. If X is a random variable which takes values x_i with probabilities represented by $P(x_i)$, then the expected value is given by $\sum_i x_i P(x_i)$.

For a set of random variables, the joint distribution of those variables can be thought of as a grid of probabilities for each possible combination of states for the set of variables. Consider another random variable S representing whether it is sunny which can take values true and false. The joint distribution of G and S is a 2×2 grid of all the possible combinations of the values of S and G and will be represented as $\mathbf{P}(G,S)$. From a joint distribution, a marginal distribution can be obtained by summing over all the possible values of one or more variables.

$$\mathbf{P}(G) = \sum_S \mathbf{P}(G,S) \tag{2.1}$$

Equation 2.1, also known as the sum rule (for example in [12]), gives a marginal probability distribution over G by summing over all the possible values of S.

A conditional probability is a probability given some other known information. Using the random variables G and S as described, $P(G = t|S = t)$ represents the probability that the grass is wet given that it is sunny. A grid of probabilities, $\mathbf{P}(G|S)$, represents the probabilities of each value of G given each value of S.

Conditional probabilities are defined in terms of unconditional probabilities as follows:

$$P(a|b) = \frac{P(a,b)}{P(b)} \tag{2.2}$$

where a and b are any events.

This equation can be re-written in the following way, which is also known as the product rule.

$$P(a,b) = P(a|b)P(b) \tag{2.3}$$

Putting some numbers to the example

$$\mathbf{P}(S) = (0.3\ 0.7)$$

$$\mathbf{P}(G|S) = \begin{array}{c c} & \begin{array}{cc} S=t & S=f \end{array} \\ \begin{array}{c} G=t \\ G=f \end{array} & \begin{pmatrix} 0.25 & 0.6 \\ 0.75 & 0.4 \end{pmatrix} \end{array}$$

This grid is a conditional probability table and represents statements such as; given that it is sunny there is a 25% chance that the grass is wet. From these two probability distributions, the joint distribution of G and S can be calculated, using Equation 2.3.

$$\mathbf{P}(G,S) = \begin{pmatrix} 0.25 * 0.3 & 0.6 * 0.7 \\ 0.75 * 0.3 & 0.4 * 0.7 \end{pmatrix} = \begin{pmatrix} 0.075 & 0.42 \\ 0.225 & 0.28 \end{pmatrix}$$

From this joint probability distribution, the marginal probability distribution over G can be obtained by summing out S. This gives

$$\mathbf{P}(G) = (0.495\ 0.505)$$

Bayes' theorem

As $P(a,b)$ is the same as $P(b,a)$, equation 2.3 can also be written as

$$P(a,b) = P(b|a)P(a) \tag{2.4}$$

Equations 2.3 and 2.4 can be combined to give Bayes' theorem

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} \tag{2.5}$$

In terms of machine learning, Bayes' theorem is often expressed as follows (for example in [13]).

$$P(hypothesis|evidence) = \frac{P(evidence|hypothesis)P(hypothesis)}{P(evidence)} \tag{2.6}$$

This gives a way of updating belief in hypotheses on receiving more observations, or evidence. The term $P(evidence)$ does not depend on the hypothesis which is being considered and so the equation can be written as

$$P(hypothesis|evidence) = \alpha P(evidence|hypothesis)P(hypothesis) \tag{2.7}$$

where α is determined in order to make all the probabilities sum to one, a process called normalisation.

2.2.3 Bayesian Networks

A Bayesian network is a graphical representation of the relationship between all the variables in a situation. Each variable becomes a node on the graph. Nodes are connected to show dependencies between the variables. Node X is joined by a directed arc to node Y if X causes Y. Each node of the graph needs to have a conditional probability table specified for the corresponding variable given its parents. Given this information, probabilistic inference can be carried out, applying evidence to update the probabilities at some nodes and computing the updated probabilities at other nodes taking the evidence into account.

If the nodes of a Bayesian network were labelled $X_1 .... X_n$ then the full joint distribution can be expressed as

$$\mathbf{P}(X_1, ..., X_n) = \prod_i \mathbf{P}(X_i | parents\ of\ X_i) \tag{2.8}$$

As Bayesian networks involve graphical representation, they are best illustrated by means of an example with the associated Bayesian network. The following example is based on one used by Jensen [14].

Holmes leaves his house and notices his lawn is wet. He wants to know whether it rained overnight or he forgot to turn off the sprinkler. Given that the lawn is wet, the probability of rain or the sprinkler being on are both increased from his initial beliefs. Holmes looks next door and notices that Watson's lawn is also wet. Holmes then reduces belief in the sprinkler having been left on.

This sort of reasoning can be illustrated numerically with a Bayesian network. According to Jensen this is carried out automatically by humans.

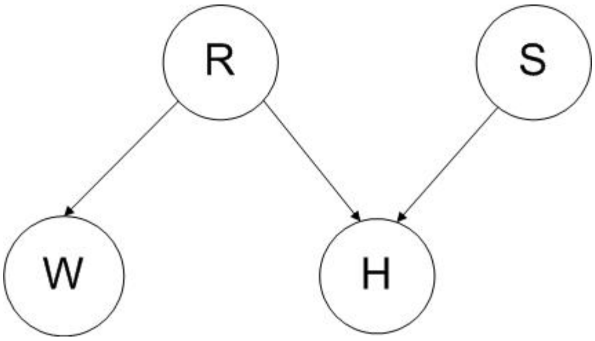


Figure 2.1: Bayesian network for the wet grass example.

In Figure 2.1, each of the variables can take just two values, true or false. R represents 'it rained last night', S 'the sprinkler was left on', H 'Holmes' lawn is wet', W 'Watson's lawn is wet'.

As stated above, each node requires a conditional probability distribution on the variable conditioned on its parents. These initial probabilities will be set as follows

$$\mathbf{P}(R) = (0.2\,0.8)$$

$$\mathbf{P}(S) = (0.1\,0.9)$$

$$\mathbf{P}(W|R) = \begin{array}{c} \\ W=t \\ W=f \end{array} \begin{array}{cc} R=t & R=f \\ \left(\begin{array}{c} 1 \\ 0 \end{array} \right. & \left. \begin{array}{c} 0.2 \\ 0.8 \end{array} \right) \end{array} \qquad \mathbf{P}(H|R,S) = \begin{array}{c} \\ S=t \\ S=f \end{array} \begin{array}{cc} R=t & R=f \\ \left(\begin{array}{c} (1,0) \\ (1,0) \end{array} \right. & \left. \begin{array}{c} (0.9,0.1) \\ (0,1) \end{array} \right) \end{array}$$

The probabilities $\mathbf{P}(R)$ and $\mathbf{P}(S)$ are prior probabilities, they are beliefs which are held before evidence has been gathered.

From these we can compute $\mathbf{P}(W,R) = \mathbf{P}(W|R)\mathbf{P}(R)$ using the definition of conditional probability as follows

$$\begin{pmatrix} 1 & 0.2 \\ 0 & 0.8 \end{pmatrix} \times \begin{pmatrix} 0.2 & 0.8 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.2 & 0.16 \\ 0 & 0.64 \end{pmatrix}$$

The symbol × is used to indicate pointwise multiplication.

Marginalising over R, that is adding both columns together to remove R, gives the prior probability of W to be

$$\mathbf{P}(W) = (0.36\,0.64)$$

In a similar way $\mathbf{P}(H,R,S) = \mathbf{P}(H|R,S)\mathbf{P}(S)P(R)$. As $\mathbf{P}(H,R,S)$ is a 3 dimensional grid, the calculations are shown separately for $H = t$ and $H = f$.

$$\begin{array}{cc} H = t & H = f \\ \begin{pmatrix} 1 & 0.9 \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{pmatrix} = \begin{pmatrix} 0.1 & 0.09 \\ 0.9 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0.1 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{pmatrix} = \begin{pmatrix} 0 & 0.01 \\ 0 & 0.9 \end{pmatrix} \\ \begin{pmatrix} 0.1 & 0.09 \\ 0.9 & 0 \end{pmatrix} \times \begin{pmatrix} 0.2 & 0.8 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.02 & 0.072 \\ 0.18 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0.01 \\ 0 & 0.9 \end{pmatrix} \times \begin{pmatrix} 0.2 & 0.8 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0 & 0.008 \\ 0 & 0.72 \end{pmatrix} \end{array}$$

Marginalising over both R and S, adding the rows and columns of the individual matrices for $H = t$ and $H = f$, gives the prior probability of H.

$$\mathbf{P}(H) = (0.272\,0.728)$$

So far, the observation that both Holmes' and Watson's lawns are both wet has not been taken into account. The joint probability distribution $\mathbf{P}(H,R,S)$ can be updated to give $\mathbf{P}^*(H,R,S) = P(H,R,S|H = t)$. The asterisk is used to indicate that this is an updated probability. This update can be done by setting all the probabilities for $H = f$, that is setting the final matrix on the right to zero. The resulting matrix is normalised by dividing each entry by the sum of the entries.

$$\mathbf{P}^*(H,R,S) = \begin{array}{c} \\ S = t \\ S = f \end{array} \begin{array}{c} \begin{array}{cc} R = t & R = f \end{array} \\ \begin{pmatrix} (0.074, 0) & (0.264, 0) \\ (0.662, 0) & (0, 0) \end{pmatrix} \end{array}$$

The notation (n_1, n_2) within the table represents $\mathbf{P}(H = t, H = f)$.

From this matrix we can marginalise to give updated probabilities for R and S given the fact that Holmes' lawn was wet.

Summing the rows gives $\mathbf{P}^*(R) = (0.736\,0.264)$. Summing the columns gives $\mathbf{P}^*(S) = (0.338\,0.661)$.

As stated in the introduction to the problem, the fact that Holmes' lawn was wet has increased the probabilities for both rain and the sprinkler having been left on. These are posterior probabilities, after evidence has been taken into account.

The new probability distribution over R is used to update $P(W,R)$

$$\mathbf{P}^*(W,R) = \mathbf{P}(W|R)\mathbf{P}^*(R)$$

$$\mathbf{P}^*(W,R) = \begin{pmatrix} 1 & 0.2 \\ 0 & 0.8 \end{pmatrix} \times \begin{pmatrix} 0.736 & 0.264 \\ 0.736 & 0.264 \end{pmatrix} = \begin{pmatrix} 0.736 & 0.0528 \\ 0 & 0.2112 \end{pmatrix}$$

It is known that Watson's lawn was also wet and so $W = t$. The entries for $W = f$ in the matrix $\mathbf{P}^*(W,R)$ can be set to zero and then the matrix normalised giving

$$\mathbf{P}^{**}(W,R|W=t,H=t) = \begin{array}{c c} & \begin{array}{cc} R=t & R=f \end{array} \\ \begin{array}{c} W=t \\ W=f \end{array} & \begin{pmatrix} 0.933 & 0.067 \\ 0 & 0 \end{pmatrix} \end{array}$$

From this, it can be seen that $\mathbf{P}^{**}(R=t) = 0.933$. The probability that it has rained has increased, from its value knowing only that Holmes' lawn is wet, on taking into account the fact that Watson's lawn is wet.

To calculate a new probability for the sprinkler having been left on,

$$\mathbf{P}^{**}(H,R,S) = \mathbf{P}^*(H,R,S)\frac{\mathbf{P}^{**}(R)}{\mathbf{P}^*(R)}$$

$$\mathbf{P}^*(W,R) = \begin{pmatrix} 0.074 & 0.264 \\ 0.662 & 0 \end{pmatrix} \times \begin{pmatrix} 0.933/0.736 & 0.067/0.264 \\ 0.933/0.736 & 0.067/0.264 \end{pmatrix} = \begin{pmatrix} 0.094 & 0.067 \\ 0.839 & 0 \end{pmatrix}$$

Marginalising gives $P(S=t) = 0.161$.

This has reduced from the probability which just takes into account the fact that Holmes' grass is wet, as expected from the problem description.

2.2.4 Temporal inference with Bayesian Networks

Bayesian networks can be used for reasoning about events which take place over a sequence of time, such networks are referred to as Dynamic Bayesian networks.

Markov models are particular examples of Dynamic Bayesian networks. A Markov chain is one in which the variable depends on its value at previous time steps.

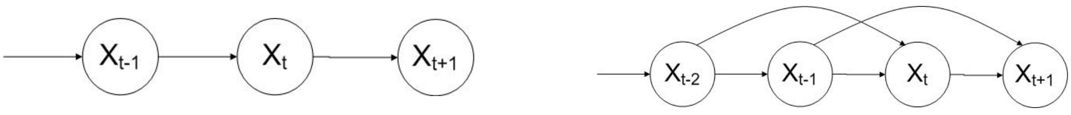


Figure 2.2: Left: First order Markov chain Right: Second order Markov chain.

Hidden Markov models are ones in which a discrete hidden variable forms a Markov chain and an observed variable depends on the hidden variable at each time point.

The Bayesian network described in the wet grass example above contains observations at only one point in time. It would be possible to consider making an observation, for example whether Watson's lawn is wet, every morning. If it is assumed that the observations are made at uniform intervals of time, the time points can be numbered. Again assuming that whether Watson's lawn is wet depends only on whether it rained overnight, the variable R is a state variable. The subscript t is used to refer to a variable at time point t. This situation can be modelled by a Hidden Markov model as shown in Figure 2.3.

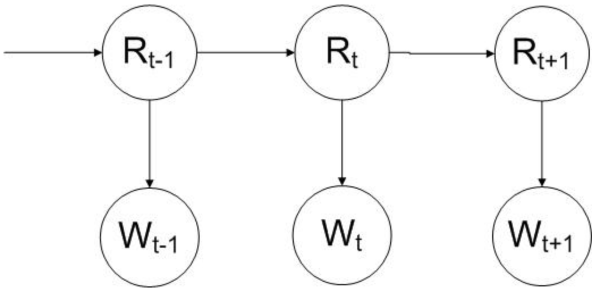


Figure 2.3: Hidden Markov Model for the wet grass example.

Hidden Markov models make a number of assumptions as given below, using the random variables from Figure 2.3.

- The state of the system depends only on the previous state.

$$\mathbf{P}(R_t|R_0,R_1....R_{t-1}) = \mathbf{P}(R_t|R_{t-1})$$

- Each outcome is independent of each other outcome given the states

$$\mathbf{P}(W_1....W_t|R_0,R_1....R_t) = \prod_i \mathbf{P}(W_t|R_0....R_t)$$

- Each outcome depends only on the state at that time, not at any other times

$$\mathbf{P}(W_t|R_0,R_1....R_t) = \mathbf{P}(W_t|R_t)$$

- The process is stationary; the transition between one state and the next does not depend on the time point.

Using these assumptions, reasoning can be carried out in a similar way to that for the Bayesian network described above. Variations on Hidden Markov models are used for the learning task, as described in Chapter 5.

Chapter 3

Psychological Experiment

3.1 Data Collection

The behavioural data from the experiment by Bland and Schaefer [1] were made available for the current report. The study had been approved by a local Ethics committee and each participant signed a consent form.

The information provided included all the details of the settings, correct response and actual response with response times for every trial for 32 participants. The data were supplied in a separate file for each participant. The files contained tabular data with column headings which I used to determine which fields to select. I used Matlab to convert the required fields to numeric format and create a single table from all the files. The details of the fields used and the conversion can be seen in Appendix E.

Although the experiment was similar to that of Behrens *et al* [2], there were a number of differences. Each participant carried out eight blocks of 120 trials, which included volatile and non-volatile blocks. In non-volatile blocks, the same underlying rule applied for all 120 trials. In volatile blocks, the underlying rule was reversed every 30 trials. The error rate was constant throughout a block with two settings, high and low. In the low error condition, 83.3% of the responses rewarded conformed to the underlying rule. For high error blocks, this was 73.3%. The underlying rule could take two different forms, the correct response when presented with a red stimulus, could be to press button one or button two. The combination of two states of volatility, two error rates and two underlying rules gave eight combinations which formed the eight blocks.

3.2 Analysis of the data

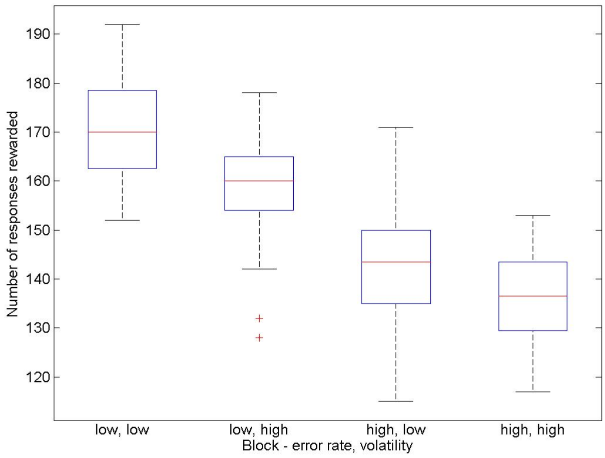


Figure 3.1: Number of rewards achieved by block type.

The data have been grouped into four different conditions by considering the error rate and volatility only, combining the trials with the rule for button one and two reversed. Figure 3.1 shows the overall behaviour of the participants in the different blocks. The boxes show the interquartile range with the line across the box giving the median. The whiskers show the range for all data which are not considered to be outliers.

The box plots show that the median number of responses rewarded was highest for the low error rate and low volatility condition and reduces with both increase in error rate and increase in volatility. This would be expected as it would be harder to identify the underlying rule when the probability of an error is high or the rule changes often. The range of rewards given is large for the low error rate, high volatility condition.

The trials which went against the current dominant rule were considered to be presented randomly, so the most profitable approach taken by the participants would be to consistently respond in favour of the current rule, that is maximising. Figure 3.2 shows how the amount of maximising relates to the rewards achieved. There is a clear linear relation between the number of maximising responses and the rewards achieved. This confirms that the most profitable strategy was maximising. There is a large spread across the participants in the number of trials in which maximising was shown.

The variation in the numbers of maximising responses according to the type of trial can be seen in Figure 3.3. It is clear that maximising was highest in the condition when both error rate and volatility were low, top left. In the high error, low volatility condition, top right, the participants appear in three separate groups rather than being closely grouped together.

The maximum number of consecutive maximising responses by each participant, shown in Figure 3.4,

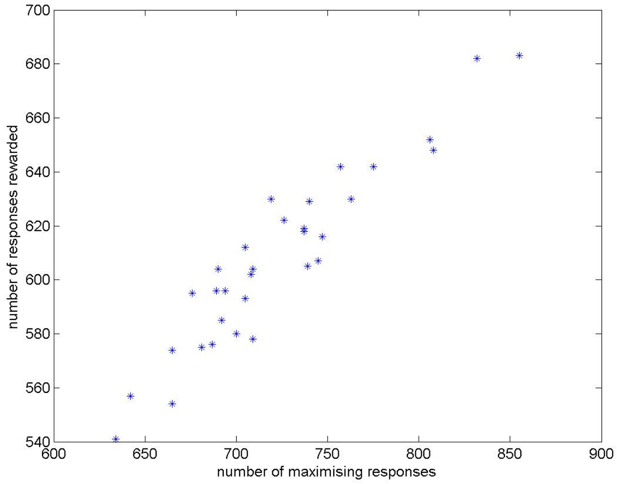


Figure 3.2: Rewards given compared to the number of maximising responses by each participant.

allows the performance to be evaluated using the measure of maximising given by Shanks *et al* [6] of maximising during at least 50 consecutive trials. Using this measure, eight participants used a maximising strategy in Bland and Schaefer's study.

As each error/volatility condition had been presented twice to each participant, I decided to consider

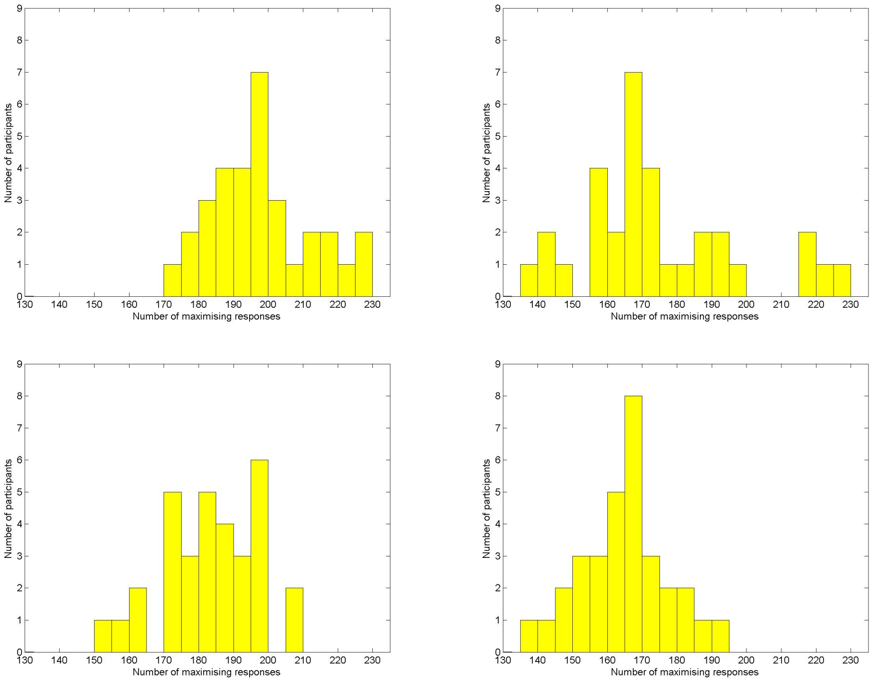


Figure 3.3: Left: Low error probability Right: High error probability Above: Low volatility Below: High volatility.

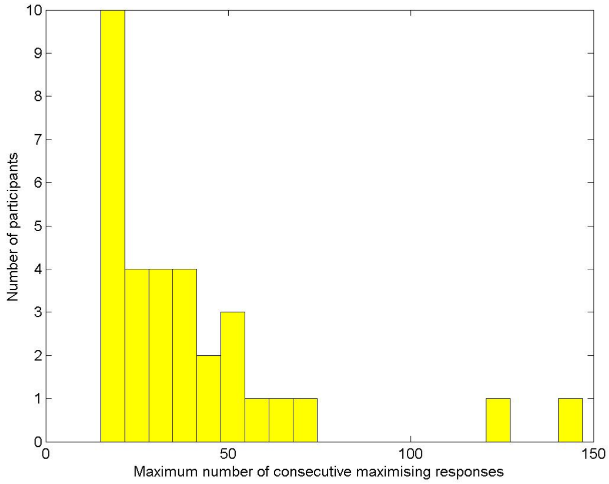


Figure 3.4: Histogram of consecutive maximising responses.

only the second time each condition had been presented. The data provided by Bland and Schaefer contained labels of whether a trial was considered to be in the first or second half of a block. For blocks with low volatility, the second half was labelled as trials 61 to 120. For high volatility blocks, the second half was labelled as the second half of each set of 30 trials.

Using data from second half of the second time a condition had been experienced, gave the most chance of observing maximising behaviour as the participant had had time to learn both the nature of the task and the current rule. This data was clustered using k-means to split the participants into two groups. The averages for each condition can be seen in Table 3.1.

The choice of two groups to use for the k-means clustering was determined by examination of the data and from descriptions of behaviour in other studies. The k-means algorithm finds the best split for the two groups. Figure 3.3 shows that there is a spread of maximising behaviour rather than two neat groups. I still felt it would be useful to separate those participants who could be identified as using a maximising strategy.

error probability volatility	low low	low high	high low	high high
Cluster 1	56.8	56.7	56.2	54.3
Cluster 2	48.6	47.1	42.4	42.5

Table 3.1: Cluster centres for maximising during second half, the second time a condition was presented.

As these figures were taken from the second half of the second presentation of a condition, they cover only 60 trials. If participants were probability matching, for the low error condition, 50 maximising responses would be expected; in the high error condition 44 maximising responses would be expected. From the table, it can be seen that cluster 1 shows an average behaviour which is close to maximising in all conditions and cluster 2 are performing a little lower than probability matching. The data forming these clusters can be seen in Figure 3.5 with cluster 2 shown with a square symbol. The six participants in the maximising cluster can be seen in the top right corner of each graph and seem to be clearly separated from the other participants.

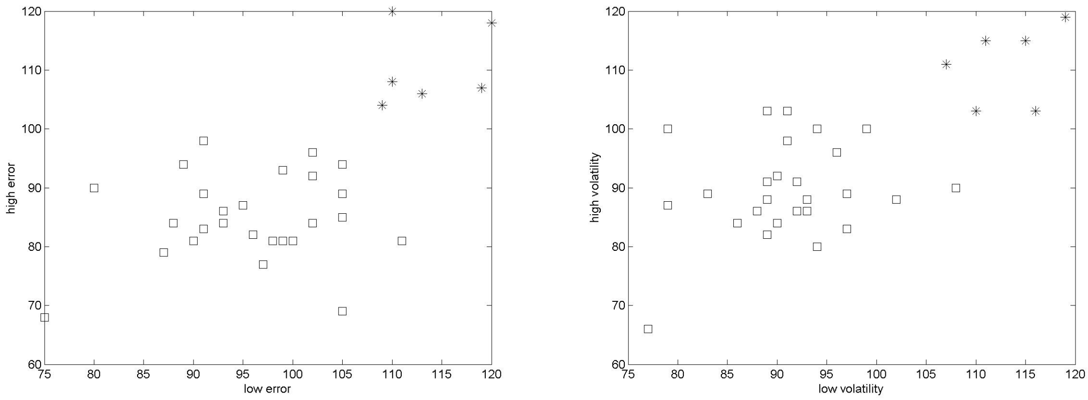


Figure 3.5: Clustering of maximising responses by error condition on the left and volatility on the right.

3.3 Discussion

3.3.1 Comparison with the work of Behrens *et al*

Bland and Schaefer looked at stimulus-response learning rather than the simple prediction of the next outcome in a sequence. This means that the participants had more to remember as they had to learn an underlying rule as well as when the correct answer was opposite to the expected rule.

Behrens *et al* used a reward value which changed on each of the 290 trials and did not deduct any points from the participants if they failed to predict correctly. The participants could see how well they were doing during the study against targets of £10 and £20. In Bland and Schaefer's study the participants saw how many points they won or lost on each trial but did not have a running total and had no information on how the points would be converted to money at the end of 960 trials. Having such a large number of trials could lead to the participants becoming bored, influencing their subjective value of the rewards given.

Behrens *et al* instructed their participants that the correct colour "depended only on the recent outcome history". Bland and Schaefer do not report a similar statement. Bland and Schaefer only allowed 1000 ms each trial for a participant to react. This aspect is not mentioned by Behrens *et al.*

3.3.2 General Comments

Bland and Schaefer presented the eight blocks in a random order to the participants. As different blocks had the same underlying rule, this meant that participants may not have seen a change in rule when a change in blocks had occurred. When comparing four different groups of data based on error rate and volatility, the participants may not have had to learn a new rule at the beginning of the block. The blocks were presented consecutively to the participants with only one break which occurred after four blocks, 480 trials. Two of the participants maximised for more than 120 consecutive trials, as shown in Figure 3.4. The randomisation of the blocks meant that only eight of the 32 participants experienced a high volatility block as their first block. The behaviour on later blocks would be influenced by the experience of earlier trials and so the different conditions may not be equivalent for comparison.

I used the data presented to the participants to examine what performance would result if the only rule used was to respond according to the rule implied from the previous trial. Only 11 of the 32 participants performed better than using this simple strategy. Just looking at the most recent outcome would not require a complicated process when the rule changed.

Considering the points raised by Shanks *et al* [6], as discussed in Section 2.1, about the set up of decision making experiments -

- The sequence of events used by Bland and Schaefer was not completely random. In blocks with high volatility, the ratio of trials fitting the current error rule exactly matched the stated error probability over each group of 30 trials. In the low volatility blocks, the ratio held exactly over 120 trials. The randomisation over 120 trials was carried out separately for each participant, leading to a different individual experience. For example, the maximum number of consecutive trials which followed the dominant rule varied from 14 to 31 between the participants.
- The rewards were displayed to the participants as a number of points only, not a monetary value.
- The participants may have been trying to find patterns in the presentation of the stimuli. In addition to an underlying rule, the participants may have believed that their responses made some difference to the outcome or that the order of presentation of blue or red stimuli affected which response would be considered correct.
- Bland and Schaefer only reported on the behaviour as a whole group. My analysis of the data, as described above, indicates that there are individual differences in behaviour. This can be seen in the level of maximising behaviour of the participants. Analysis of the behaviour of the group as a whole gives only limited insight into the learning process and is not appropriate for drawing conclusions about the presence or absence of Bayesian reasoning.

Chapter 4

Modelling

4.1 Data generation

The data provided by Bland and Schaefer [1], described in Chapter 3, gave outcomes for 960 trials for each of 32 participants. If a Bayesian learner performs well on this data, it might be that the learner has been tuned to this specific data but would not perform well with a different set of data with a similar underlying structure. To ensure that any differences found between models related to the learning task, I chose to use a much larger volume of data, for this reason I created simulated data.

The study by Bland and Schaefer has additional complexity compared to other experiments on learning rules including that of Behrens *et al* [2], as discussed in Section 2.1. Bland and Schaefer expected the participants to learn that there was an underlying rule to the association of a colour with a button, a stimulus response rule. This would give two variables which were observed by the participants, the colour seen and the rule which was rewarded.

Application of the model used by Behrens *et al*, required a single sequence of observations. From the data of Bland and Schaefer, this was taken to be the rule which was rewarded, without taking account of the colour shown. This was a binary choice, associating one button with a colour presented assuming that the other colour was then associated with the other button. I used the symbols ‘1’ and ‘2’ to represent the two choices. I converted the outcomes to a sequence of ‘1’s and ‘2’s which just indicate which rule was rewarded on each trial. I could then treat the data as identical to predicting the next outcome in a sequence and create new data following this pattern.

To see whether significant differences in human behaviour between high/low error/volatility levels reported by Bland and Schaefer were also found in the decisions of Bayesian learners, I needed to use the same definitions of these terms, including the number of trials in a block, where a block is a group of trials having the same experimental condition. Low error blocks had a reward rate of 83.3%, that is, 83.3% of outcomes match the dominant rule, equivalent to an error rate of 16.7%. In high error conditions, 73.3% of outcomes matched the dominant rule. Volatile blocks had the rule reversed every 30 trials, stable blocks had a single rule for all 120 trials.

I created 2000 blocks of 120 trials for each of the four combinations of error and volatility levels. This data was stored separately by error rate with each stable block followed immediately by a volatile block. A Matlab random number generator was used to determine the outcome at each trial according to the error rates above. Each trial outcome was independent of each other, just using the required reward rate rather than sticking to the exact proportion in each block of 30. I validated the data by counting the numbers of ‘1’s and ‘2’s across rows and columns and the proportions were found to be correct.

4.2 General modelling decisions

Behrens *et al* used a continuous variable, r, to represent the probability that the next observed outcome would be blue, where $0 \leq r \leq 1$. Instead of allowing continuous variation in r, I replaced r by a discrete random variable R representing the current belief about the probability of observing a ‘1’. This random variable, R, can take N possible values, denoted by r_i for $0 \leq i \leq n$ where $n = N - 1$. Each r_i represents a probability of observing a ‘1’ and so for each i, $0 \leq r_i \leq 1$. The N points are taken to be evenly spaced in the [0,1] interval, that is $r_i = i\Delta r$ where $\Delta r = \frac{1}{N-1}$. These values r_i can be written as a vector $(r_0, ..., r_n)$, represented by $\mathbf{r}$. A probability can be assigned to each r_i, giving a vector of probabilities $(P(R = r_0), ..., P(R = r_n))$ this vector will be represented as $\mathbf{P}(R)$. As R can only take these n possible values, the sum of the probabilities for each of these values has to be one, $\sum_i P(R = r_i) = 1$.

For example, if N is taken to be 5, then $\mathbf{r} = (0, 0.25, 0.5, 0.75, 1)$. Using this example, $\mathbf{P}(R) = (0, 0, 0, 1, 0)$ expresses the belief that the probability of observing a ’1’ is 0.75 and that there is no uncertainty about this belief. A probability distribution $\mathbf{P}(R) = (0, 0, 0.25, 0.5, 0.25)$ expresses the belief that the probability of observing a ‘1’ is between 0.5 and 1 with the most likely value being 0.75 but with some uncertainty about the actual value. In order to use Bayes’ theorem, a prior distribution over R is needed, this represents the belief that the outcome will be ‘1’ before any observations have been made. The prior probability distribution is taken to be uniform, that is an equal probability is given to each possible value of R. In the case where $n = 5$, this would be given by $\mathbf{P}(R) = (0.2, 0.2, 0.2, 0.2, 0.2)$.

The random variable X refers to the observation and so can take values of '1' or '2'. The subscript t is used with a variable to refer to a particular trial in a sequence. An observation which has been made at trial t will be denoted x_t. The first observation will be considered to be made at time $t = 1$. Time $t = 0$ will be used for the prior probabilities which are assumed before any observations are made.

Using Bayes' theorem requires a value to be given to the probability of making a given observation for each possible r_i, that is $P(x|r_i)$, where x is the observation. As each r_i is defined as the probability of observing a '1', for each r_i, if a '1' is observed, then $P(x|r_i) = r_i$. In the vector notation $\mathbf{P}(x|R) = \mathbf{r}$. As the only possible outcomes are '1' or '2', then for each r_i the probability of observing a '2' is $1 - r_i$. Therefore if a '2' is observed, $P(x|r_i) = 1 - r_i$.

The probability distribution over R expresses uncertainty about the actual probability of observing a '1'. In order to make a prediction as to the next observation in sequence, the mean or expected value of R is used. Griffiths *et al* [15] state that the mean is the probability which should be used to predict the next outcome in the case of tossing a biased coin. Behrens *et al* also use the mean to make a prediction for the next outcome. The mean of the probability distribution of R is given by $\sum_i r_i P(R = r_i)$. This can be computed by taking the scalar product of vectors $\mathbf{r}$ and $\mathbf{P}(R)$.

Using discrete points meant that the probability distributions, including joint probability distributions where necessary, could be stored as grids. In addition, the number of points used could be reduced down to two, giving each variable two possible states, high and low. The inference process could be carried out manually for a few steps and used to validate that the model was behaving exactly as required. For experimentation, I needed to choose a number of points into which to split the interval [0,1]. There needed to be enough points to allow a reasonable approximation of the continuous model used by Behrens *et al* and also to represent probability distributions graphically. In addition there needed to be not so many points that the probability associated with each individual value r_i became too small to carry out computation. The number of points also needed to be small enough to allow matrix computation in reasonable time. I chose to use 49 points. Matlab was used as a programming language due to its ability to manipulate data stored as vectors and matrices and to create graphical output from that data.

4.3 Simple application of Bayes' theorem

Bayes' theorem as stated in Equation 2.7 in Section 2.2 is as follows.

$$P(hypothesis|evidence) = \alpha P(evidence|hypothesis)P(hypothesis)$$

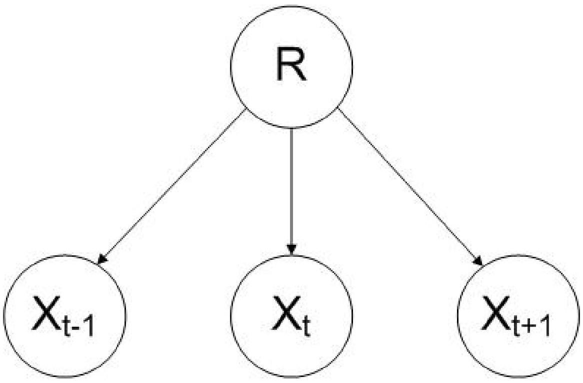


Figure 4.1: Simple Bayesian Model

Applying this to the model shown in Figure 4.1 and explicitly stating the prior knowledge at time t, that is all the outcomes up to time $t-1$ gives

$$\mathbf{P}(R|x_1....x_t) = \alpha \mathbf{P}(x_t|R)\mathbf{P}(R|x_1......x_{t-1})$$

where $\mathbf{P}(R|x_1......x_{t-1})$ is the prior probability distribution for R which is updated based on the observed outcome at trial t by multiplying by $\mathbf{P}(x_t|R)$. This requires multiplication by $\mathbf{r}$ if a '1' is observed, as described in Section 4.2 above. This multiplication needs to apply separately for each possible value r_i of R. Storing the probabilities as vectors requires pointwise multiplication to be used. This gives the posterior probability distribution for R, $\mathbf{P}(r|x_1....x_t)$. The mean of this posterior probability distribution is used to make a prediction for the outcome at trial $t+1$. The mean is taken as described above. A '1' is predicted for the next outcome if the mean is greater than 0.5. The posterior probability distribution of R at trial t becomes the prior probability distribution at trial $t+1$. To begin the process, an initial probability distribution for R is required, this is taken as a uniform distribution.

The Bayesian reasoning process used here will also be used in further, more complex, models which are described in Chapter 5. The performance of the models will be evaluated in Chapter 6.

Chapter 5

Further Models

5.1 General points

Following the simple Bayesian model described in Section 4.3, a hierarchy of models was created by successively increasing the number of variables or parameters. The construction of these models built on the decisions described in Section 4.2. Each of the following models is a variation on a Hidden Markov Model. In these models, the process is described by one or more discrete hidden variables which form a first order Markov chain. These models are variations on Hidden Markov models as they do not all have just one hidden variable. The models depend on the assumptions described in Section 2.2.4. Making a first-order Markov assumption, does not limit the performance too strongly as, according to Russell and Norvig [10], an increase in order of a Hidden Markov model can always be produced by a model with an increased number of hidden variables.

In addition to the variable, r, representing the probability of the next outcome taking a particular value, Behrens *et al* [2] introduced variable, v, volatility which controls the rate at which r changes, and parameter k which controls the rate at which v changes. As r represents a probability, it has to vary between 0 and 1 only. Behrens *et al* do not specify the range in which v and k are allowed to vary. Discrete equivalents of these variables are used, with grids to store the data, requiring end values to be set. The discrete variables are represented by R for reward rate, V for volatility and K as a parameter, replacing r, v and k respectively as used by Behrens *et al.* I chose to allow each variable to range between 0 and 1.

With Hidden Markov models, transition matrices need to be defined to specify the change in each hidden variable from one time point to the next. Following Behrens *et al*, I used a beta distribution to determine the new probability distribution for R_{t+1} based on the probability distribution of R_t. The transition matrices were determined in order to match the qualitative descriptions by Behrens *et al* of the required behaviour over the interval.

A beta distribution is a probability distribution which takes two parameters a and b. The values of the parameters determine the shape of the curve. The mean of the distribution is given by $a/(a+b)$. The sum of the two parameters, $a+b$, is a measure of the spread of the distribution. Figure 5.1 shows increasing values of $a+b$, labelled as sum, each for three different values of the mean of the distribution. The beta distribution for a given value of $a+b$ and a mean of one minus those shown, would be the curve formed by reflection in the line $x=0.5$. Note that in Figure 5.1, the y-axis is not identical for each value of $a+b$.

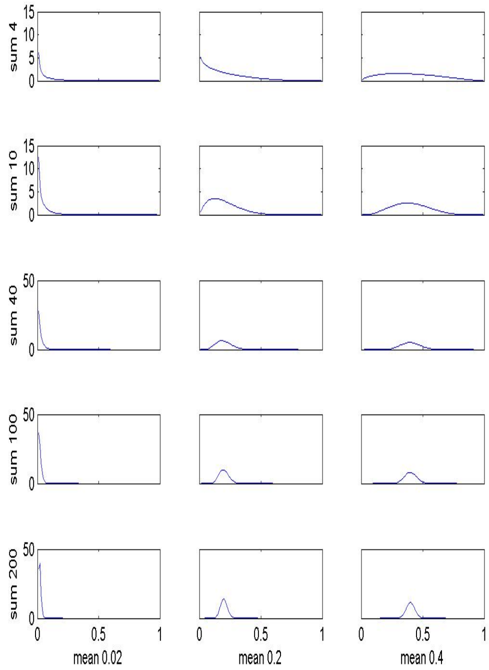


Figure 5.1: Beta distribution for different values of the sum of the two parameters.

The beta distribution is used by Behrens *et al* such that the probability density function for r is centred around the mean of r from the previous trial. The use of a large spread, that is a small value for $a+b$, allows r to change greatly from one trial to the next. Using a large value of $a+b$ gives a distribution which has a sharp peak and can force r to only change a little between trials. This behaviour would be useful in a stable situation to prevent a change in prediction when presented with occasional outcomes which are opposite to the general rule. In this way, $a+b$ can be used as a measure of confidence in r.

The beta distribution is only defined in the interval (0,1), that is, it is not defined at the end points 0 and 1. The beta distribution has behaviour which means that the value approaches infinity towards the end points of this range. This can be seen in Figure 5.1 for a mean of 0.02. This led me to truncate the interval to [0.02,0.98] to avoid computational errors arising from very large values in some cells of a matrix and very small values in others.

One of the reasons that a beta distribution is often used is that the area under its curve is always 1 so fits the requirements of a probability density function for a continuous variable. Using discrete variables and a slightly truncated range, the distribution had to be re-normalised. This truncating of the range and normalisation meant that the expected value of R_{t+1} would tend towards 0.5. This truncation would have more effect when the expected value of R_t was near the edges of its range. As the only information required to make predictions for outcomes was to compare the mean of R_{t+1} to 0.5, the truncation would not affect the predictions.

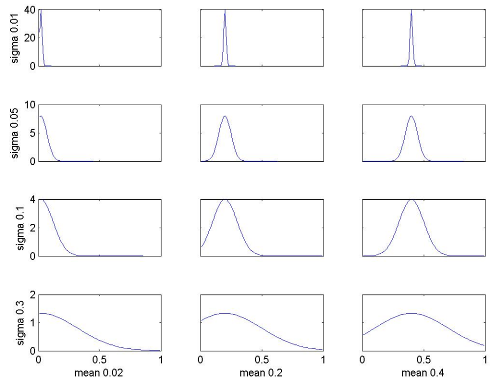


Figure 5.2: Normal distribution for different values of sigma.

A normal distribution calculated with the mean of each possible value v_i of V is used for the transition from V_t to V_{t+1}, again following the work of Behrens *et al.* The normal distributions shown in Figure 5.2 shows increasing values of the standard deviation, sigma. The normal distribution is not limited to the range (0,1) but was truncated to that range. Again the probabilities had to be normalised.

5.2 Hidden Markov Model variations

5.2.1 Model with one hidden variable

A graphical representation of this model is shown in Figure 5.3. This model is a Hidden Markov model as it has one discrete hidden variable. The parameter R used in the simple application of Bayes' theorem described in section 4.3 is replaced by a hidden variable, R_t. In the simple application of Bayes' theorem, the belief in the probability distribution for R is updated directly by application of Bayes' theorem on receiving new evidence. The expected value of R changes very slowly from one trial to the next. In this Hidden Markov model, the change in the probability distribution for R_t from one time step to the next is determined by a transition matrix which is applied before each outcome,

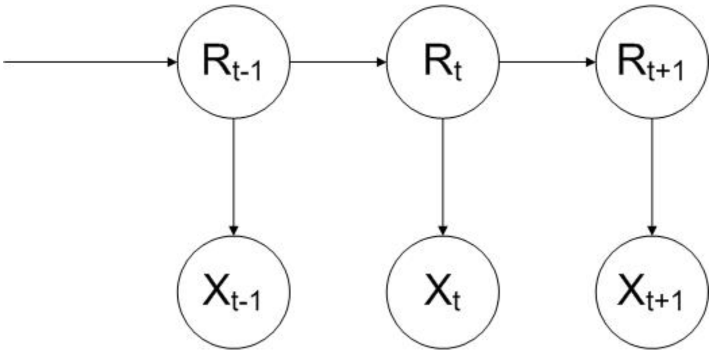


Figure 5.3: Model with one hidden variable.

x_t, is taken into account. Using a transition matrix allows the rate of change of R_t to be controlled. A transition matrix which allows only a small change from R_t to R_{t+1} gives a model which behaves in the same way as the simple application of Bayes' theorem. Using a transition matrix which allows a large change from one time to the next can create a model which makes predictions in accordance with only the most recently observed outcome.

The equation for probabilistic inference using this model can be derived as follows

$$\mathbf{P}(R_{t+1}|x_1,....,x_{t+1}) = \alpha\mathbf{P}(R_{t+1},x_1,....,x_{t+1})$$

from the definition of conditional probability

$$= \alpha\mathbf{P}(x_{t+1}|R_{t+1})\mathbf{P}(R_{t+1},x_1,....,x_t)$$

using Bayes law and the fact that x_t depends only on R_t

$$= \alpha\mathbf{P}(x_{t+1}|R_{t+1})\sum_i \mathbf{P}(R_{t+1}|R_t = r_i)P(R_t = r_i,x_1,....,x_t)$$

as R_{t+1} only depends on R_t and having taken the sum over all the possible values of R_t

$$= \alpha\mathbf{P}(x_{t+1}|R_{t+1})\sum_i \mathbf{P}(R_{t+1}|R_t = r_i)P(R_t = r_i|x_1,....,x_t)$$

from the definition of conditional probability

In each of the above, α represents a constant used to normalise the probabilities so that they sum to 1.

The inference results in a probability distribution $\mathbf{P}(R_{t+1}|x_1,....,x_{t+1})$ over R_{t+1} which is stored as a vector of probabilities for each possible value of R_{t+1}.

The transition matrix $\mathbf{P}(R_{t+1}|R_t)$ is a probability distribution for R_{t+1} for each possible value r_i of R_t. This represents the transition from one time step to the next and is unchanged through the sequence. This transition is stored as a 2-dimensional matrix. For each value r_i of R_t, I determined the probability distribution of R_{t+1} from a beta distribution with the mean at r_i and a constant value for the sum of the two parameters throughout an experiment. This gives a constant transition matrix for an experiment, corresponding to one row from Figure 5.1. Experiments carried out with different transition matrices are described in Section 6.1.1.

The final line of the inference process contains the term, $P(R_t = r_i|x_1, ..., x_t)$. This is obtained from the result of the probabilistic inference from the previous time step. This gives a recursive procedure.

The multiplication and summation over r_i can be carried out by matrix multiplication. This summation gives a probability distribution for R_{t+1} given all the observations up to time t. The probability distribution over R_{t+1} is used in the same way as that over R in the simple application of Bayes' theorem as described in Section 4.3. The mean is taken to make the prediction for next outcome. The observation at time $t+1$ is taken into account by multiplication by $\mathbf{P}(x_{t+1}|R_{i+1})$.

To begin the inference process, the prior probability distribution of R_t before any observations have been made, $\mathbf{P}(R_0)$ is taken to be a uniform distribution.

5.2.2 Model with one hidden variable and one parameter

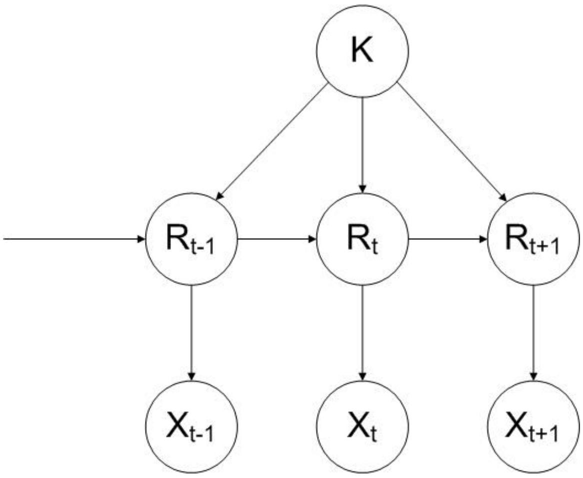


Figure 5.4: Model with one hidden variable and one parameter.

The model represented in Figure 5.4 is an extension of the model described in the previous section by adding an additional parameter, K, which controls the transition from R_t to R_{t+1}. The parameter K allows the model to have more than one transition matrix. The parameter K can be seen as a measure of confidence in the current probability distribution over R_t. The application of Bayes' theorem allows the probability distribution over K to change during an experiment. The parameter K is discrete and takes values k_i where for each i, $0 \leq k_i \leq 1$.

As the parameter K would be replaced by the variable for volatility in more complex models, the probability distribution over R_t was made to change more quickly for high values of k_i and more slowly for low values of k_i. Given the limited details provided by Behrens *et al* and the modelling decisions described in Section 5.1, I used the qualitative descriptions of the behaviour to determine the transition matrices. Using plots of beta distributions such as Figure 5.1, I chose to allow the sum of the parameters to vary between 10 and 100. A value of 10 would allow R_t to change greatly between trials and 100 would only allow a small change between one trial and the next.

Behrens *et al* used an exponential function to control how $a+b$ depended on volatility. Looking at Figure 5.1, small changes make a big difference to the shape of the distribution when $a+b$ is around 10, but larger changes in $a+b$ are needed to make a noticeable difference when $a+b$ is around 100. This suggests that it is appropriate to use a function which showed exponential growth. I chose to set $a+b$ to be 10^{2-k_i} allowing $a+b$ to range between 100 and 10 as required.

The equation for inference for this model is

$$\mathbf{P}(R_{t+1}, K | x_1, ..., x_{t+1}) = \alpha \mathbf{P}(x_{t+1} | R_{t+1}) \sum_i \mathbf{P}(R_{t+1} | R_t = r_i, K) \mathbf{P}(R_t, K, x_1, ..., x_t)$$

The transition matrix $\mathbf{P}(R_{t+1}|R_t, K)$ is stored as a 3D grid. This is formed from a 2D transition matrix $\mathbf{P}(R_{t+1}|R_t)$ for each possible value k_i of K. Each of these is equivalent to a single transition matrix used in the simpler Hidden Markov model above, described in Section 5.2.1. The computation within the sum is carried out for each value k_i of K to give a joint probability distribution over R_{t+1} and K. A distribution on R_{t+1} alone is formed by summing out over K. The expected value of the distribution over R_{t+1} can then be used to make the prediction as in the previous models.

The joint distribution over R_{t+1} and K is updated by multiplying by $\mathbf{P}(x_{t+1}|R_{t+1})$ to take into account the observation at time $t+1$. This is the same as in the previous model but for each k_i of K.

The prior probability distribution over R_0 and K was varied to look at the effect of differing priors, this experimentation is described in Section 6.1.2.

5.2.3 Model with two hidden variables

The complexity of the model represented in Figure 5.5 is increased compared to the previous one. Instead of a parameter K, there is a hidden variable V_t which represents the volatility of the system. Having two hidden variables requires two transition matrices to be defined. The transition from V_t to V_{t+1}, $\mathbf{P}(V_{t+1}|V_t)$, is a 2D matrix. As discussed in Section 5.1, this transition matrix is determined by a normal distribution. In this model, a constant value of the standard deviation is used. The transition

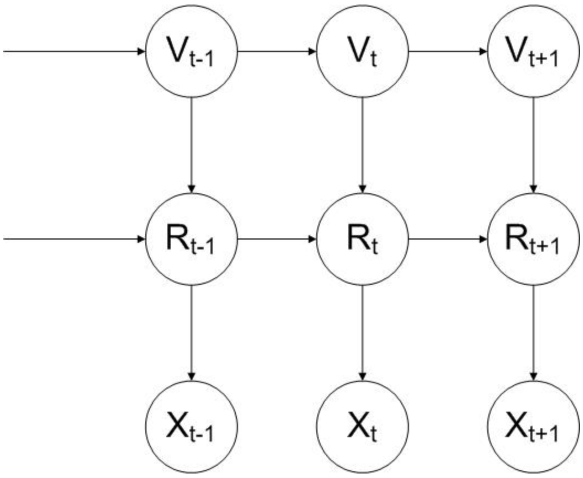


Figure 5.5: Model with two hidden variables.

from R_t to R_{t+1} requires 3 dimensions and is the same as that used in the previous model with K replaced by V_t.

The differences between this model and the previous one mirror those between the first Hidden Markov model considered with only one hidden variable, described in Section 5.2.1 and the simple application of Bayes' theorem described in Section 4.3. The hidden variable V_t can be made to change more rapidly than the parameter K in the previous model by selection of an appropriate transition matrix for V_t.

The equation for inference for this model is

$$\mathbf{P}(R_{t+1},V_{t+1}|x_1,....,x_{t+1}) = \\ \alpha\mathbf{P}(x_{t+1}|R_{t+1})\sum_i(\mathbf{P}(R_{t+1}|R_t=r_i,V_{t+1})\sum_j\mathbf{P}(V_{t+1}|V_t=v_j)\mathbf{P}(R_t,V_t=v_j|x_1,....,x_t))$$

This requires a nested summation which is carried out from right to left. The first summation can be carried out by matrix multiplication on two 2D matrices and gives a joint probability distribution over V_{t+1} and R_t. The second summation is then identical to that in the previous model but replacing K by V_{t+1}. The prediction of the next outcome and updating with the next result are carried out in the same way as in the previous model.

The prior probability distribution over R_0 and V_0 at the start of the experiment is taken to be uniform.

5.2.4 Model with two hidden variables and a parameter

A discrete equivalent of the model used by Behrens *et al* is shown in Figure 5.6. This matches the graphical representation given by Behrens *et al* who claim that their model is optimal in tracking a

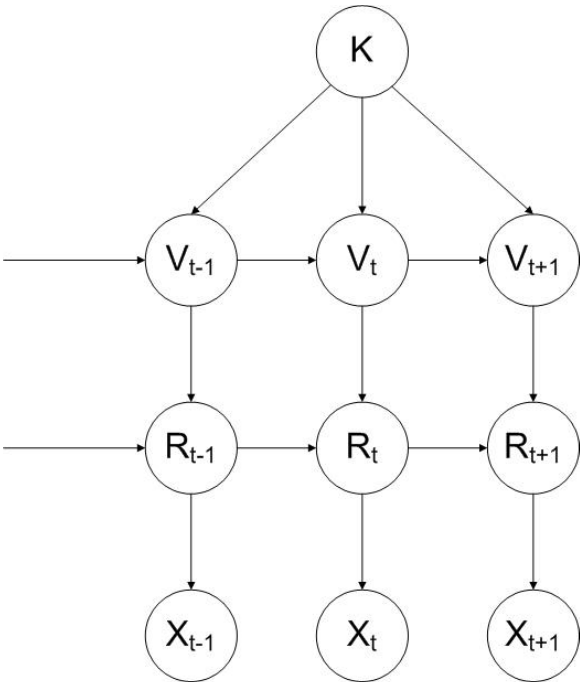


Figure 5.6: Model with two hidden variables and one parameter

varying reward. This is the most complex model, in terms of the number of parameters and variables, considered in this work and will be referred to as the complex model.

The parameter K in the model removes the requirement in the previous model to select a fixed transition matrix for V_t, in the same way as the model with one hidden variable and one parameter, described in Section 5.2.2, has a set of transition matrices for R_t.

As with the model with one hidden variable and one parameter, I needed to decide how the transition matrices were determined from the values k_i of the parameter K. Behrens *et al* used k to determine the standard deviation of a normal distribution. The effect of varying the standard deviation for given mean values is shown in Figure 5.2. As in the use by Behrens *et al*, high values of k_i should cause V_t to vary greatly between trials and low values of k_i allow little variation. I chose to allow the standard deviation to vary between approximately 0.01 and 0.6.

The equation for inference for this model is

$$\begin{aligned}&\mathbf{P}(R_{t+1}, V_{t+1}, K|x_1, ..., x_{t+1}) = \\ &\qquad \alpha\mathbf{P}(x_{t+1}|R_{t+1})\sum_i(\mathbf{P}(R_{t+1}|R_t = r_i, V_{t+1})\sum_j \mathbf{P}(V_{t+1}|V_t = v_j, K)\mathbf{P}(R_t, V_t = v_j, K|x_1, ..., x_t))\end{aligned}$$

The transition matrix for V is now 3 dimensional as one 2D matrix is required for each value k_i of K. The result of the inference, which is stored between trials is a joint probability distribution over R_{t+1},

V_{t+1} and K and so requires a 3D grid. The summations can be carried out as in the previous model but replicated for each value k_i of K.

In order to make a prediction, the joint probability distribution needs to be marginalised over both V_{t+1} and K to give a probability distribution over R_{t+1} alone from which the expected value can be calculated.

To start the inference process, a uniform joint probability distribution over R_0, V_0 and K is taken, as in the work of Behrens *et al.*

5.3 Validation of the models

5.3.1 Comparison with hand calculations

The computational models were constructed so that the number of discrete points into which to split the (0,1) interval was set by a parameter. This allowed the models to be run with the (0,1) interval split into just two discrete points. The transition matrices were set separately in this case but the main computations were carried out in exactly the same way as for the full experiments. The calculations were carried out manually for the first few iterations of the process and the resulting joint probability distributions compared to those obtained using the program. In each case, the joint probability distributions matched thus validating the programming.

The process for the first three iterations for the most complex model can be found in Appendix F.

5.3.2 Number of points to use in discretisation

As explained in Section 4.2, the (0,1) interval was split into 49 discrete points for most of the experimentation using the computational models. The model with one hidden variable and one parameter was used to test whether the number of points made a big difference to performance of the model. The model was presented with all the test data and the proportion of maximising responses was used as a measure of performance.

number of points	25	50	100	150
maximising responses (%)	89.39	89.36	89.25	89.21

Table 5.1: Percentage of responses which were maximising according to the number of points in the interval.

It can be seen from Table 5.1 that varying the number of points in the interval (0,1) made a difference of less than 0.2% to the performance of the model. It was assumed that the differences would be similarly small for the other models.

5.3.3 Comparison with plots published by Behrens *et al*

Behrens *et al* included plots of joint probability distributions between pairs of variables at different stages in the experiment as shown in Figure 5.7. These probability distributions are obtained by marginalisation of the joint distribution on r, v and k. The variables r, v and k all have the same interpretation as the discrete random variables R, V and K used in the complex model described in Section 5.2.4. The top two graphs show that the system is stable after 120 trials with a constant reward rate of 75%. The volatility is low and the low value for k means that the system is not likely to change. In the middle two plots, 15 trials after a change to a reward rate of 80% in favour of the other outcome, there is still a high probability for the previous values of r and v but the probability distribution is more spread over a range of values. The bottom two plots show that the probability distribution for r is clustered around 0.2, which is an accurate reflection of the new reward rate. The parameter k is clustered around a value which is clearly higher than above, “ensuring that it would react faster to any future change in reward rate”.

Although my experiments were not identical to those of Behrens *et al*, I hoped to be able to produce plots to demonstrate that my model was behaving in a qualitatively similar way. For Behrens *et al*, a volatile block had switches in reward rate every 30 or 40 trials, whereas my data, following the paradigm used by Bland and Schaefer, had a switch in reward rate every 30 trials. I had determined transition matrices for R and V based on the qualitative description by Behrens *et al* after imposing fixed ranges to V and K. These decisions were described in Sections 5.1, 5.2.2 and 5.2.4.

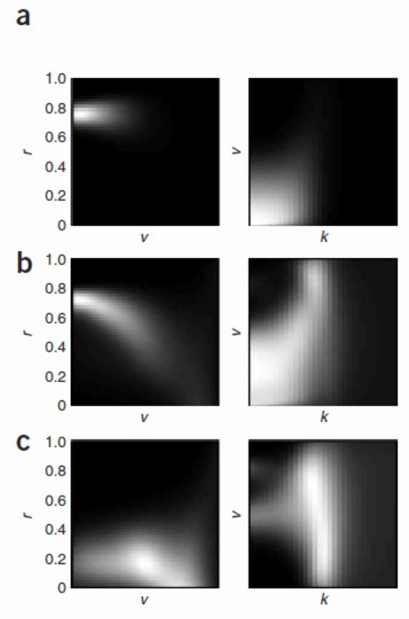


Figure 5.7: Probability distributions of variables at different stages of the experiment a) After 120 trials in a stable environment. b) Fifteen trials after a rule change. c) After another 25 trials with the same reward rate as at point b. Taken from Behrens *et al* [2].

I discovered that the probability distributions for the complex model vary depending on the individual sequence of data used, even though the parameters for generation of the data were identical. Two plots from data with a constant reward rate of 83.3% with a switch in direction after 120 trials, are shown in Figure 5.8.

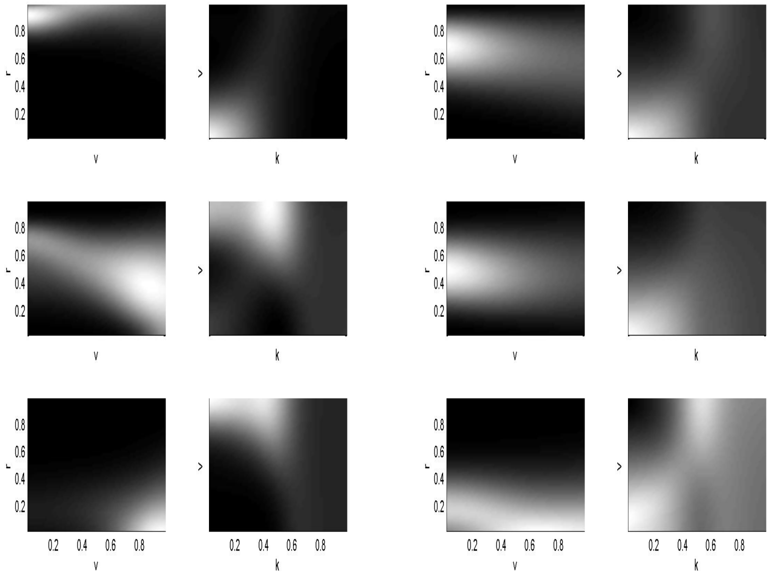


Figure 5.8: Probability distributions. Top: after 120 trials. Middle: after 130 trials. Bottom after 140 trials. Left and right show separate sequences of data.

These two plots are both different from each other and from the plot shown in Figure 5.7. Behrens *et al* do not report whether each participant and the Bayesian learner were tested against a single sequence of data or whether, like Bland and Schaefer, they generated new data fitting the parameters each time the experiment was carried out. To give a view of the behaviour of the variables which depended less on the individual data sequence, I chose to average over 50 sets of data. Behrens *et al* used a higher error rate for the stable period than for the volatile period. To approximate this, I used the data previously generated, with reward rates of 73.3% and 83.3%. The results can be seen in Figure 5.9.

After 120 trials, the probability distributions of the variables look similar to those of Behrens *et al*, with the probability distribution for R clustered around the actual reward rate of 73.3%. The probability distributions for V and K are both focussed on low values, indicating a stable system. Ten trials after a switch in reward, R is clustered around a lower value, there is some spread in the values for V, but the probability distribution for K is still tightly clustered at the lowest end of its range. The probability distribution for R changed much more quickly than shown by Behrens *et al.* A quicker change in R would allow a model to react more quickly to a switch in reward. However, a quick change in the expected value of R would not allow a model to maximise well in a stable environment with a high error rate as it would change its prediction in response to a short sequence of outcomes.

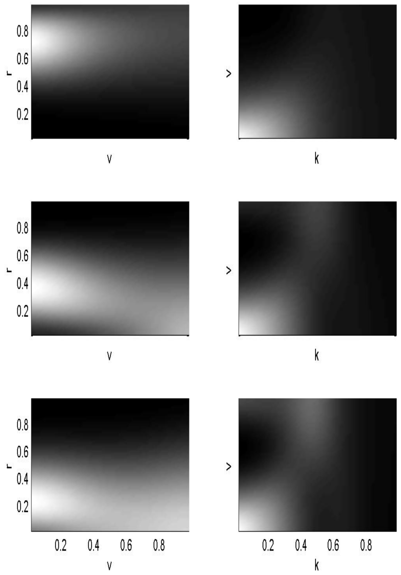


Figure 5.9: Probability distributions. Top: after 120 trials. Middle: after 130 trials. Bottom after 140 trials.

The rapid change in the probability distribution for R shown in Figure 5.9 has taken place without a noticeable change in the probability distribution over K.

I wanted to replicate the probability distributions given by Behrens *et al* more closely. This would demonstrate that the discrete model was a reasonable implementation of the work of Behrens *et al*. The rate at which the probability distributions of the variables can change from trial to trial is controlled by the transition matrices. The random variables R and V each have sets of transition matrices which depend on V and K respectively.

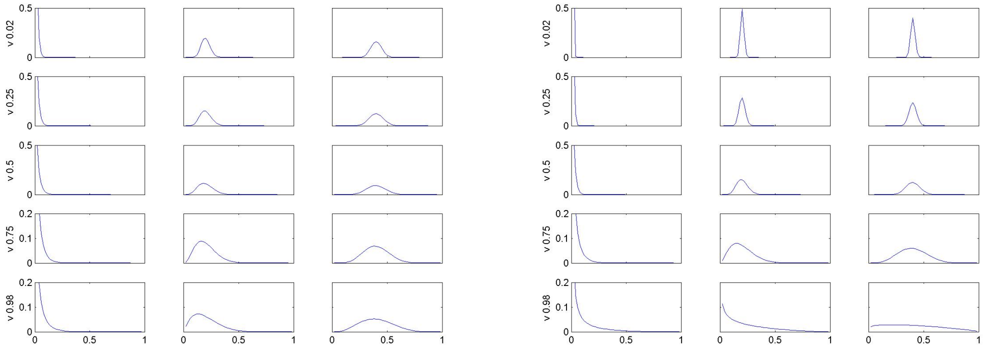


Figure 5.10: Beta distributions used for transition of R. Left: old. Right: new.

To make a variable change more slowly, the transition matrix had to have a lower probability of taking a value different to the current one. A transition matrix which does not allow the probability distribution of a variable to change at all will be a diagonal matrix. The closer a transition matrix resembles a diagonal matrix, the smaller the change which is allowed in the variable. The transition matrices are built from beta and normal distributions as explained in Section 5.1. A slow change would be represented by a sharper spike.

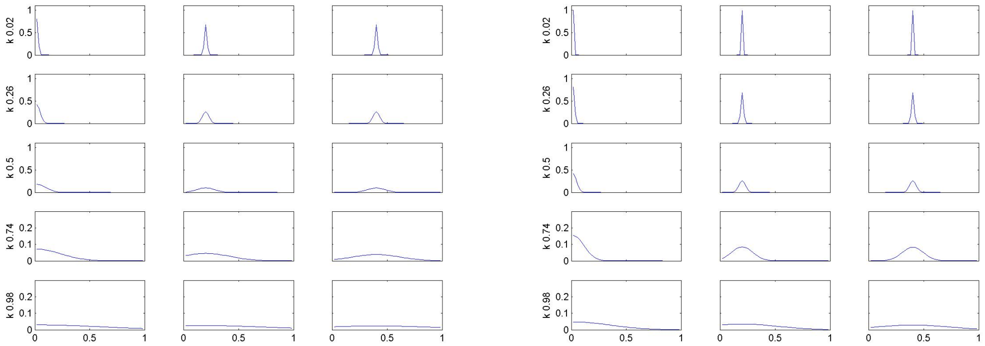


Figure 5.11: Normal distributions used for transition of V. Left: old. Right: new.

The parameters for the beta distribution were amended to give a wider range of behaviour across the range of values for V. For low values v_i of V, the peaks were made sharper and for high v_i the peaks spread more. This can be seen in the beta distributions used for the previous and new transition matrices for R in Figure 5.10. The value v_i of V is increasing between 0.02 and 0.98 down the rows. The transition matrix for V was changed in a similar way, with the transition matrix depending on the value k_i of K. Plots of the previous and new functions used can be seen in Figure 5.11. Note that the y-axis is not the same for each row.

$$\begin{pmatrix} 0.9944 & 0.0056 & 0 & 0 & 0 & \dots & 0 \\ 0.0056 & 0.9888 & 0.0056 & 0 & 0 & \dots & 0 \\ 0 & 0.0056 & 0.9888 & 0.0056 & 0 & \dots & 0 \\ 0 & 0 & 0.0056 & 0.9888 & 0.0056 & \dots & 0 \\ 0 & 0 & 0 & 0.0056 & 0.9888 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0.9944 \end{pmatrix}$$

$$\begin{pmatrix} 0.7441 & 0.2468 & 0.0090 & 0 & 0 & \dots & 0 \\ 0.1980 & 0.5968 & 0.1980 & 0.0072 & 0 & \dots & 0 \\ 0.0072 & 0.1966 & 0.5925 & 0.1966 & 0.0072 & \dots & 0 \\ 0 & 0.0072 & 0.1965 & 0.5925 & 0.1966 & \dots & 0 \\ 0 & 0 & 0.0072 & 0.1965 & 0.5925 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0.7441 \end{pmatrix}$$

Figure 5.12: Sections of transition matrices for V Top: $K = 0.02$ Bottom $K = 0.3$.

There was a limit to how sharp the peaks could be made due to the choice of 49 discrete values to split the range [0.02,0.98]. If I tried to make the peaks too sharp then a probability could be assigned to a value but no probability of moving from that value. I ensured this by checking the resulting transition matrix did not contain any values of exactly 1. Transition matrices for V are shown in Figure 5.12. The matrix above shows the transition matrix for $K = 0.02$, this is close to a diagonal matrix. For each v_i of V, there is probability of 0.99 of remaining at that value. The lower matrix is that used for $K = 0.3$, now the probability of keeping the same value is 0.59 with a spread each side.

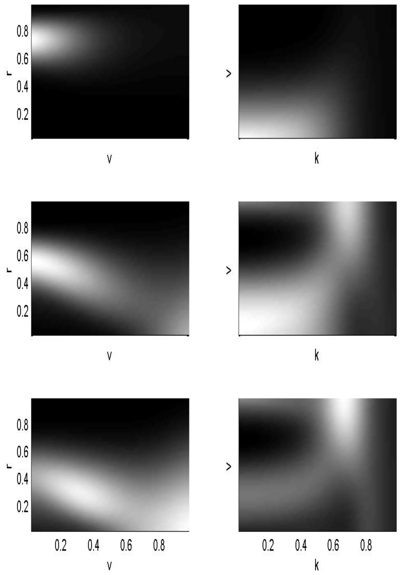


Figure 5.13: Probability distributions after changes to transition matrices. Top: after 120 trials. Middle: after 130 trials. Bottom after 140 trials.

The probability distributions resulting from the changes to the transition matrices can be seen in Figure 5.13. These probability distributions appear much more like those of Behrens *et al*, Figure 5.7. Much more of the range of possible values of K is being used and the probability of V taking a large value is higher than in Figure 5.9. The mean value of R still seems to change more quickly than in the work of Behrens *et al*. I believe that the qualitative similarity to the work of Behrens *et al* shows that an appropriate replication of their work has been carried out.

Chapter 6

Experiments

6.1 Experiments with the models

The performances of different models are compared by considering the proportion of the responses which match the current underlying rule, referred to as maximising as described in Section 2.1. A high percentage of maximising would indicate that the learner has a good estimate of the reward rate and is not switching too quickly when faced with outcomes opposite to the current rule. A model which performs well has a good balance between ignoring or responding to outcomes which are opposite to the current rule to perform well in both stable and volatile conditions.

Using the previously generated data, as described in Section 4.1, a model was run for two consecutive blocks each time. The blocks were presented stable first then volatile and, separately, volatile first then stable keeping the error rate the same throughout a run. Behrens *et al* [2] also reversed the presentation of the stable and volatile blocks to prevent ordering effects. The results were averaged over all the blocks with the same conditions and used to compare different models.

6.1.1 Varying the transition matrix

In the simplest Hidden Markov model tested, with one hidden variable, described in Section 5.2.1, the transition from the probability distribution on R_t to that for R_{t+1} is fixed throughout an experiment. The overall ability of the model to track changes in reward depends on the transition function chosen. If the transition process always makes the expected value of R_{t+1} very close to that of R_t then the model will not be able to respond quickly to a switch in reward. However, if the transition allows the expected value of R_{t+1} to vary greatly from that of R_t, then the model will effectively just take account of the most recent outcome and not take stable phases into account.

Using a beta distribution to define the transition matrix, as described in Section 5.1, I varied the sum of the two parameters a and b and examined the overall behaviour with each different value chosen.

Figure 6.1 shows that the behaviour is qualitatively different for the different volatility levels. Maximising behaviour is clearly higher in low volatility than in high volatility. For both volatility settings, maximising is higher for low error conditions than for high error. This difference is not as great as the difference in maximising between the volatility levels. In low volatility, maximising behaviour seems to peak where $a+b$ takes a value of about 100. This gives a transition which has quite a narrow spread around the previous value r_i of R_t and so does not switch its prediction when a few trials go against the dominant rule. For high volatility, maximising behaviour decreases as $a+b$ increases. When the reward switches, a high value of $a+b$ means that the expected value for R_t only changes by a small amount each time and so it takes many trials to respond to a switch.

The value of $a+b$ which gave the best overall performance across all the conditions was 20 which produced maximising responses in 89.45% of trials.

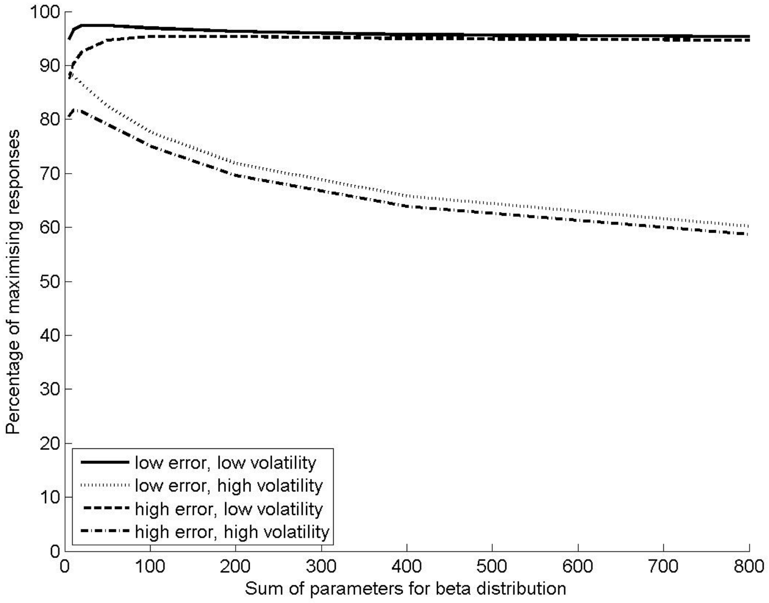


Figure 6.1: Maximising in different conditions when varying the beta distribution used for the transition matrix for R_t.

6.1.2 Varying the prior

In the model with one hidden variable and one parameter, the parameter K, can be thought of as a measure of confidence in the current probability distribution for R_t, as described in Section 5.2.2. A prior joint probability distribution over R_0 and K is needed in order to begin the inference process. The prior probability distribution was originally set to be uniform in both R_0 and K. This gave an equal probability to each possible combination of values of R_0 and K. At the beginning of the experiment, it would seem more plausible to set values for K which express a lack of confidence in the probability distribution over R_0 so I decided to try different priors which expressed increasing levels of uncertainty. The prior joint probability distribution over R_0 and K is stored in a 2D grid. A beta distribution was used, with different parameters, to express different levels of uncertainty. If the rows of the grid represent the possible values of R_0, then each row can be set to a probability distribution over K.

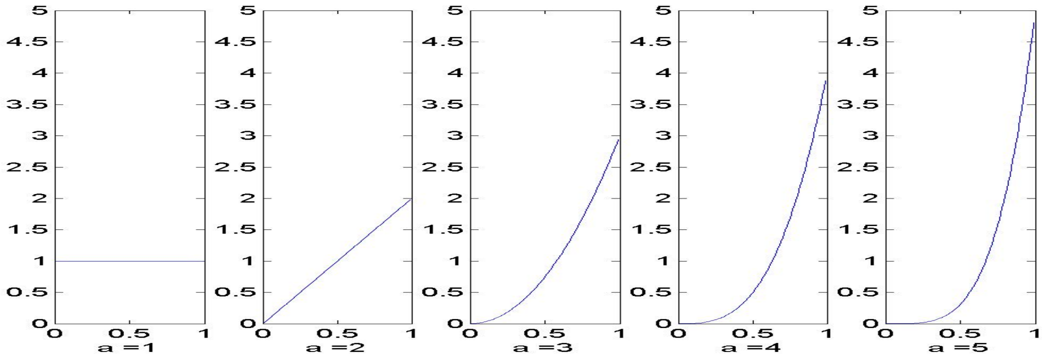


Figure 6.2: Examples of beta distributions used to vary the prior of the parameter K

Figure 6.2 shows the effect of varying the first parameter, a, in a beta distribution. As a increases, the mean of the distribution moves towards 1. A higher value of the mean of K would express a higher level of uncertainty about the probability distribution of R_0. This would allow R_t to change more quickly from one trial to the next, as described in Section 5.2.2. The effect of the initial prior distribution would be most noticeable during the first few trials of an experiment.

For each value of a chosen, each row in the joint distribution was set to the values calculated from a beta distribution with that value of a. Every row in the joint distribution was identical, the joint distribution was then normalised so that the probabilities summed to 1. If the joint distribution was marginalised to get a distribution over R only, it would be uniform. The distribution over K alone would resemble one of the beta distributions shown in Figure 6.2.

Tests were carried out using all of the generated data as previously described. Figure 6.3 shows the results of using some different priors over K set up in the way described above. The overall performance can be seen to vary only slightly with the different priors, with a very slight advantage in

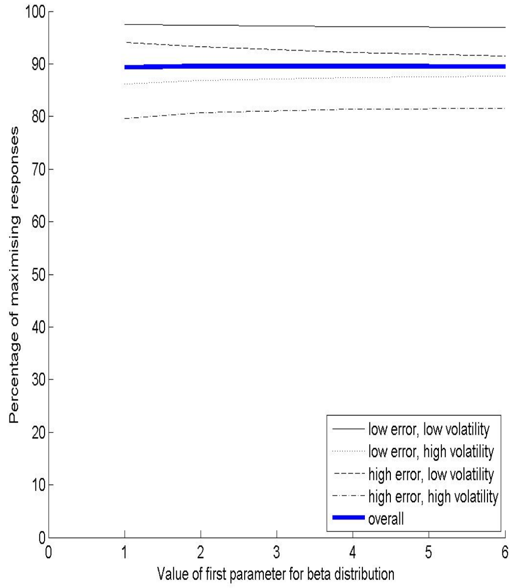


Figure 6.3: Maximising in different conditions when varying the prior used for the parameter K.

setting a to 2. Performance is reduced slightly in low volatility conditions for increasing a although these are averaged over runs with volatile blocks first and second. For high volatility, increasing a improves performance.

6.1.3 Transition matrices in the complex model

The most complex model tested, described in Section 5.2.4, was an implementation of the model used by Behrens *et al*. As described in Section 5.3.3, the transition matrices were amended to cause the model to behave more like that of Behrens *et al* in terms of the probability distributions of the variables, Figure 5.7. I wished to discover how much difference these changes made to the performance in terms of maximising over the different conditions.

volatility	error	old	new
low	low	97.58	97.54
low	high	93.93	94.80
high	low	85.84	85.83
high	high	80.37	77.02
Total		89.43	88.80

Table 6.1: Percentage of responses which were maximising by the two versions of the complex model.

In the results, shown in Table 6.1, 'old' refers to the transition matrices used to produce the probability distributions shown in Figure 5.9 and 'new' refers to the model used for Figure 5.13, which is claimed to be closer to that used by Behrens *et al*. Overall, the new settings gave a slightly worse performance.

The difference in performance can be seen to mostly arise from differences in performance in high error conditions.

Behrens *et al* stated that the probability distributions of the variables after 160 trials as shown in Figure 5.7 were "ensuring that it would react faster to any future change in reward rate". This claim was tested by looking at the rate of adaptation to a new reward rate after a switch.

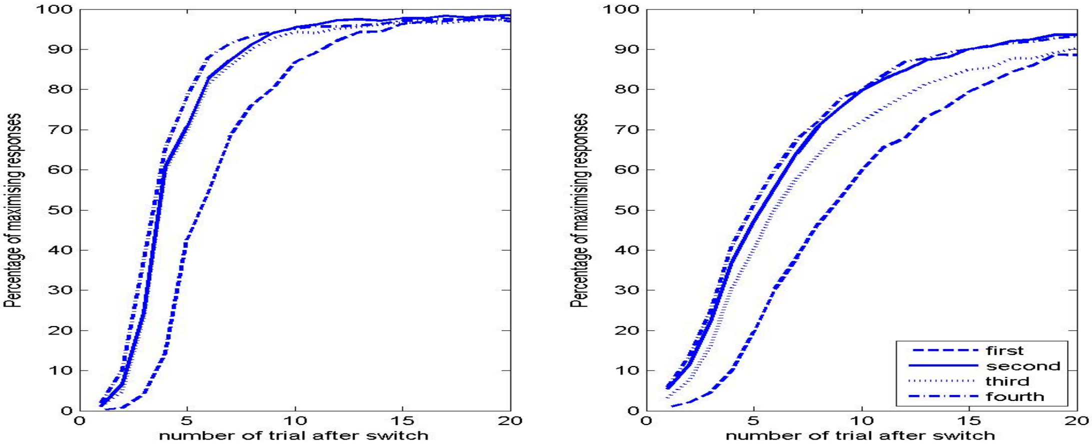


Figure 6.4: Comparing the timescale of responses to switch in reward for the complex model. Left: Low error. Right: High error.

For the test data used, the first switch occurred after 120 trials, the second 30 trials later and then at intervals of 30 trials. After each switch, the error rate remained the same but the opposite outcome became dominant. Figure 6.4 shows the percentage of trials which showed maximising behaviour by the number of the trial following a switch in reward. This was created from the performance over all the generated data. In both the low error and high error conditions, the performance after the first switch is qualitatively lower than after the subsequent switches as predicted by Behrens *et al*. Comparing the two graphs, response to a switch is clearly quicker in low error conditions than in high error. In addition, the rate of change to the new reward rate is quicker in low error conditions, shown by the steepness of the curve. In high error conditions, the steady level reached is lower than in high error.

6.2 Comparisons of the models

6.2.1 Overall performance

Each model was tested against the generated data and the results are shown in Table 6.2. Where different versions of a model existed, from the experiments described above, the performance of the version with the best overall maximising percentage is given. The variations on Hidden Markov

volatility	error	S	1H	1H, 1P	2H	2H, 1P
low	low	97.40	97.35	97.14	97.64	97.58
low	high	95.28	92.46	92.11	94.13	93.93
high	low	51.21	86.53	87.43	86.08	85.84
high	high	51.57	81.45	81.37	80.00	80.37
Total		73.87	89.45	89.51	89.46	89.43

Table 6.2: Percentage of maximising responses by each model in each condition. Models are referred to by the number of hidden variables H and the number of parameters P. S refers to the simple Bayesian model.

models are labelled by the number of hidden variables, H, and parameters, P, and ordered in order of description of the model in Chapter 5. Using this abbreviation, the complex model is labelled '2H, 1P'.

It can be seen from the table that the simple Bayesian model has a much lower performance than the models based on Hidden Markov models. The simple Bayesian model does not respond, in terms of predictions made, to a change to a volatile environment. After 120 trials in a stable environment, a sequence of 30 outcomes in the opposite direction does not change the belief in the reward rate enough to switch to the opposite prediction.

Each of the variations on a Hidden Markov model has very similar overall performance. The performance of the model with one hidden variable and one parameter, labelled '1H, 1P' and described in Section 5.2.2, is slightly better than the other models. Compared to the complex model, this simpler model shows higher maximising in high volatility and a lower rate of maximising in low volatility conditions. If participants were using Bayesian reasoning as suggested by Behrens *et al*, there is no reason to suggest a complex model when a simpler model performs almost as well.

Behrens *et al* claim that their model is optimal in tracking a varying reward. They also produce and alternative model which assumes that, rather than varying continuously, the reward is constant and then jumps to a new value. They show that there is very little difference in performance between their two models. If the complex model were close to optimal in performance, then, as the performance of the models is very similar, they would all be close to optimal.

6.2.2 Response to a switch in reward

As discussed in Section 6.1.3, the complex model, built to replicate the work of Behrens *et al*, showed a quicker response to a change in reward on the second switch than the first. The performance of the other variations on Hidden Markov models was also examined in the trials following a switch in reward. In the case of the complex model, the transition matrices which are referred to as 'new' in Section 5.3.3 are used. Figure 6.5 shows the performance following first switch after 120 trials, left, in a stable low error condition and then after a second switch 30 trials later, right. After the first switch,

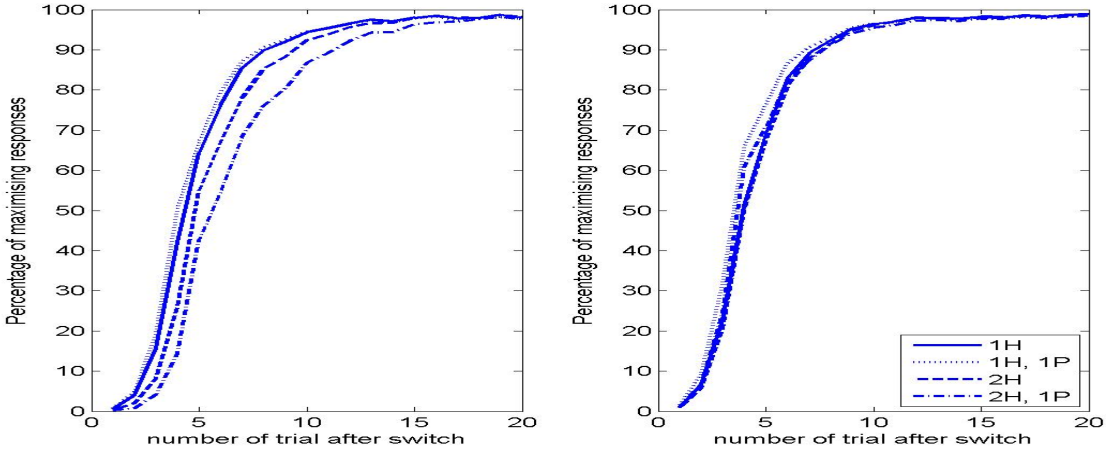


Figure 6.5: Comparison of models to switch, low error condition. Left: First switch after 120 trials in a stable state. Right: second switch.

the performance of the two most complex models, with two hidden variables is qualitatively lower than the other models until a plateau is reached.

In Figure 6.6, the same information is shown for the high error conditions. In this case, the difference between the models is more noticeable and the more complex models have lower performance than the simpler models even after the second switch.

This analysis again shows that the complex model is not as good at maximising rewards in a volatile environment as other models.

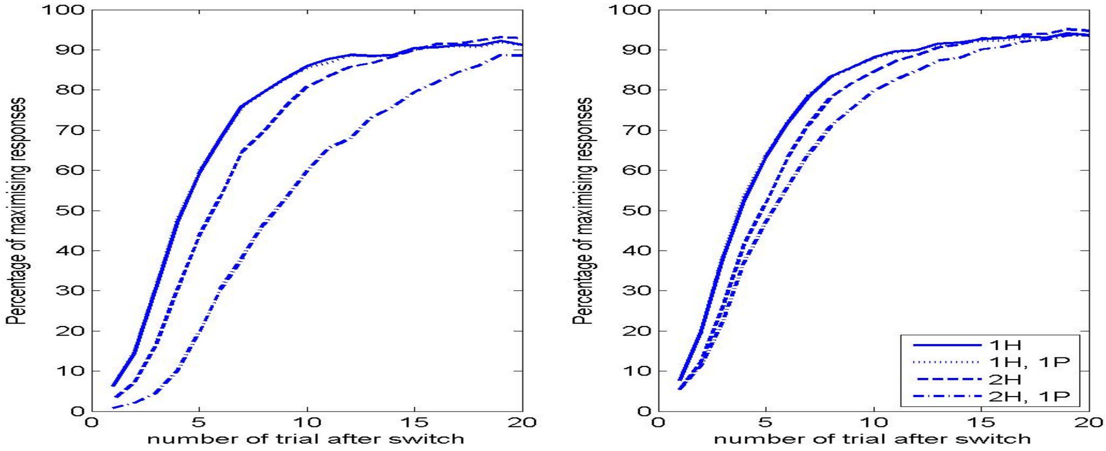


Figure 6.6: Comparison of models to switch, high error condition. Left: First switch after 120 trials in a stable state. Right: second switch.

6.3 Comparison with human learning

6.3.1 Overall performance

volatility	error	human	machine
low	low	82.59	97.79
low	high	73.14	94.92
high	low	76.60	87.58
high	high	68.93	80.38
Total		75.32	90.17

Table 6.3: Percentage of responses which were maximising by humans and machines.

The complex model with the transition matrices set to replicate the work of Behrens *et al* was used to make predictions against the data provided by Bland and Schaefer. The prior probability distribution was reset to be uniform over R_0, V_0 and K for each participants' data. The average percentage of maximising by experimental condition for the 32 participants and the model is shown in Table 6.3.

There is a very clear difference in performance between the human participants and the model with the model outperforming the humans in every condition. As has already been observed in Chapter 3, the participants do not exhibit maximising behaviour, so the model had been expected to perform better than the humans. The table shows the average behaviour. It was also the case that the model had a higher proportion of maximising than any of the individual participants.

The Bayesian model does not have to learn that the optimal behaviour, given a reward rate, is to predict according to the dominant outcome, this is a feature built in to the model. Humans have to both develop an internal representation of the changing situation and select an appropriate method for choosing a response.

6.3.2 Response to a switch in reward

One comparison between the different experimental conditions which was used by Bland and Schaefer was to look at the behaviour over successive trials following a change in reward. The performance of humans and a Bayesian model is shown in Figure 6.7. The complex model was used in this comparison. Both humans and the model reach a plateau in performance but there is a marked difference in the level of maximising reached at that plateau. For both humans and the model, the plateau is at a lower level of maximising in the high error than the low error conditions and is also reached after a larger number of trials in high error than low error conditions.

In addition to considering the participants as a single group, I decided to look at the behaviour of the six participants previously identified as possibly using a maximising strategy in Section 3.2 compared to the rest of the participants, referred to as non-maximisers. The results were also plotted against

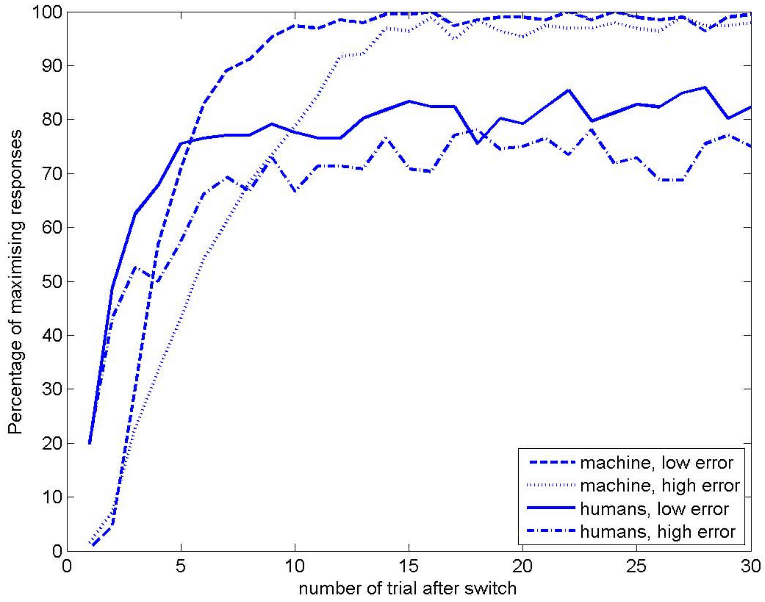


Figure 6.7: Maximising behaviour in trials after switch in reward.

a baseline model which just copies the previous outcome. A model which just copies the previous outcome reaches its plateau on the second trial after the switch. The plateau reached has the proportion of maximising equal to the reward rate, also called probability matching as described in Chapter 2. It was also observed in Chapter 3 that, for many of the participants, maximising responses were no more frequent than probability matching.

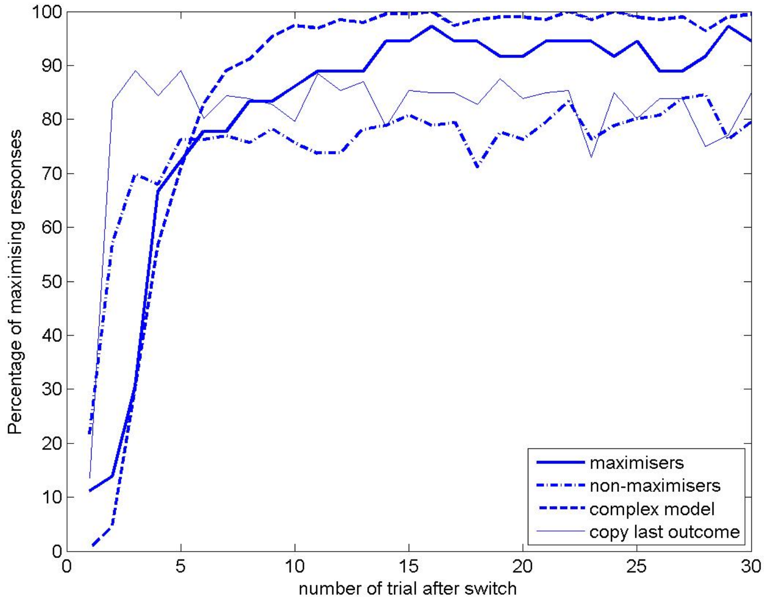


Figure 6.8: Maximising behaviour in trials after switch in reward in low error conditions.

For low error conditions, maximising behaviour following a switch in reward can be seen in Figure 6.8. The participants identified as maximisers take more trials to reach their plateau and are outperformed by the non-maximisers over the first five trials following a switch, as is the complex model. The maximisers' behaviour would suit situations in which there is a long period of stability between changes; the non-maximisers' behaviour could suit situations which change very frequently. The non-maximisers do not reach a plateau as quickly as the simple, copy last outcome model, suggesting that a more complex process is being used to determine their responses than simple copying.

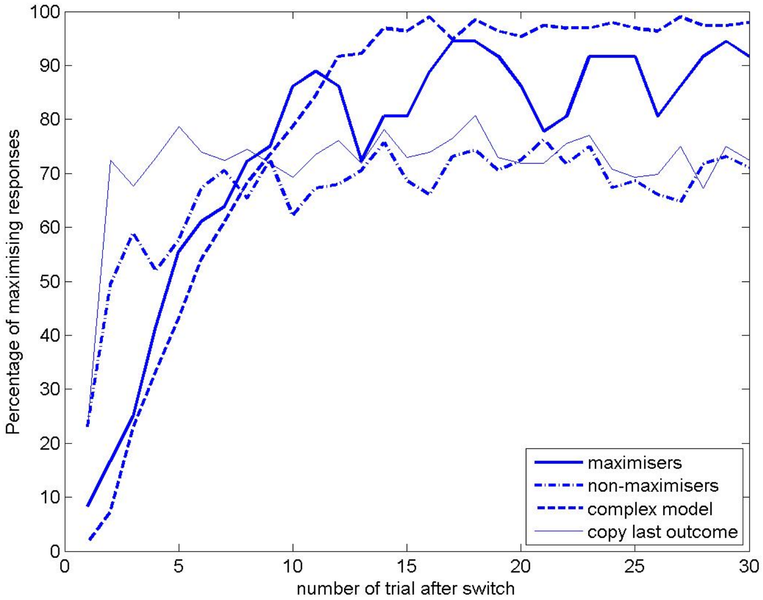


Figure 6.9: Maximising behaviour in trials after switch in reward in high error conditions.

The same details for the high error condition are shown in Figure 6.9. In this case, there are large fluctuations in the performance of the participants identified as maximisers. This is probably due to fluctuations in the actual experimental data and the fact that a group of only six participants is too few to consider overall trends.

6.3.3 Behaviour after a loss

Bland and Schaefer identified a significant difference in human behaviour in terms of whether the behaviour switched after losing on the previous trial. A switch was defined as having occurred as when a participant's response had changed in terms of maximising responses, which they termed 'adequate' responses. A switch was considered to have occurred if an adequate response was followed by an inadequate one or vice-versa. Also, a switch was deemed to have occurred when a participant failed to respond in an individual trial. Bland and Schaefer considered the experimental condition of the previous trial and whether the participant had received feedback of being correct or wrong, a loss.

The complex model was presented with the same data and the responses analysed in the same way. The model did not have the possibility of failing to respond.

For the human data, Bland and Schaefer found that in high volatility conditions, there was a significantly higher percentage of switches following a loss in low error than in high error conditions. This can be seen in Table 6.4 as the difference between 61% and 56%. There was no significant difference in the low volatility conditions, both error conditions showing switches in behaviour after 58% of losses. The difference in behaviour can be described intuitively as; in high volatility conditions, participants are expecting a switch to occur. In low error conditions when expecting a switch, a loss is seen as an indication of a rule change and so the participant changes their response. In the high error condition, more losses are expected as the error rate is high so the participant is less likely to change their behaviour in response to an individual loss.

volatility	error	human	model
high	high	56	30
high	low	61	26
low	high	58	15
low	low	58	11

Table 6.4: Percentage of switches after a loss in the different conditions.

The results of the human participants compared to the complex model are shown in Table 6.4. It can be seen that the percentages for switching after a loss are much lower for the computational model than for the human participants. No statistical tests have been carried out on the figures for the model, but it is clear that the difference found in the human data has not been found. From the table, 26% is less than 30% not greater. The percentage of switches after a loss in low error is lower than in high error for the computational model. This applies in both conditions of volatility. For the model, the percentage of switches after a loss is higher in high volatility than in low volatility for each error condition.

Chapter 7

Discussion

The psychological study by Bland and Schaefer [1] involved participants identifying a button with a colour, that is learning a stimulus-response rule. Participants were expected to recognise that there were both random fluctuations in the rewards given and changes to the underlying stimulus response rule. For the participants, the process involved both updating an internal representation of the underlying structure of the situation and then selecting an action, a button press. It was clear from examination of the data, as discussed in Chapter 3, that, considering the participants as a single group, the responses were not optimal. An optimal strategy for responding would involve always choosing the dominant outcome once the current rule had been inferred, referred to as maximising. Human behaviour often reaches a plateau in which the rate of maximising approximates the error rate, referred to as probability matching. This has been found in many studies, as discussed in Chapter 2.

One explanation for probability matching is given by Tversky and Kahneman [16] who explain that people often misunderstand the nature of chance events. People expect that even in a short sequence of events, the ratio of outcomes should match that expected. This leads people to believe that, in a stochastic binary outcome situation, such as tossing a coin, after a run of one outcome, there is a higher chance of the other outcome to balance things out. In fact, each outcome is independent of each other.

The participants who did not use a maximising strategy, the probability matchers, could have been using Bayesian reasoning to estimate the reward but it is difficult to try to separate the learning from the response. Probability matching can also be achieved by following a simple strategy like copying the previous outcome.

Computational models fitting a hierarchy of complexity were built to perform the learning task. Each model used Bayesian reasoning to infer the reward rate during the experiment, these models are described in Chapters 4 and 5. The most complex model was shown to be a discrete implementation of the model of Behrens *et al* by qualitative analysis of probability distributions as discussed in Section 5.3.3. Although Behrens *et al* only use a continuous formulation to describe their model, they also refer to “a 3-dimensional grid representing the joint distribution p(r,v,k) which is stored between trials”. Reference to a grid implies that at some point there was discretisation in their work too.

The computational models were tested against generated data. The performance of each of the variations of Hidden Markov models was very similar as discussed in Section 6.2. Even if it is proposed that humans use Bayesian reasoning, it is not necessary to suggest a model with as many variables as that used by Behrens *et al*. The model of Behrens *et al* could be used as a generative model, that is a structure with which to generate the sequence to present to the participants. Although a generative model can be used to make predictions, other models may also be used for this task.

Behrens *et al* point out that the Bayesian learner is not designed for a particular reward/volatility structure. It would have been interesting to vary the structure of the experiment more significantly from those of Behrens *et al* or Bland and Schaefer. The error rate could have been changed along with the lengths of the blocks before switching. It might be found that, compared to a model with just one hidden variable and a parameter, the additional complexity in the model of Behrens *et al* would actually give better performance in some situations.

Behrens *et al* compared human responses to two different models, reinforcement learning and the Bayesian model. In reinforcement learning, agents amend their estimate of the reward by updating it after a loss in the appropriate direction by an amount called a learning rate. Each participant has an individual learning rate which Behrens *et al* estimated using Bayesian techniques. Behrens *et al* stated that the Bayesian model had “no free parameters to fit to the subject data” and was “a significantly better predictor of subject decisions than a reinforcement learning model” and took this as evidence for Bayesian learning.

The Bayesian model used by Behrens *et al* does have parameters, these are built into the structure and transition matrices. These parameters were considered to be the same for each participant. One other option would be to consider the structure to be a parameter and use model selection techniques to find the model best fitting a participant’s responses.

This work has not made any numerical comparisons between various models to measure the level of similarity to the human responses. This decision was taken when it was found that the human behaviour did not fit into a single group, with only six participants potentially maximising, as discussed in Chapter 3. The original hypothesis that humans learn in a Bayesian way could no longer be directly tested. Psychological studies consider people as a group in order to establish whether findings are significant. For Bland and Schaefer, grouping participants was essential for analysis of the EEG data.

Other studies have tried to find the most suitable model to fit to participants' responses. Steyvers *et al* [17] looked at the responses of 451 participants to a task with four choices. They were specifically looking at individual differences in response patterns. They considered four different strategies of differing complexity. They introduced an additional parameter, *w*, to each model for use in generation of responses, this parameter controlled the rate of guessing. They used Bayesian model selection methods to determine which of the strategies best fitted the participants' data. They found that 47% of participants best fitted a win-stay strategy, 22% success ratio, 30% optimal and 1% just guessing. Having a large number of participants allowed them to consider the participants in sub-groups. The strategies of success ratio and optimal would both be the equivalent to maximising in a two option situation with no switches in reward.

In the current work, the computational models used were deterministic in nature, the outcomes were used to update probabilities in a fixed way. Following the work of Behrens *et al*, the prediction for the next outcome was that which was most likely given the current estimate of the reward. Behrens *et al* only commented on participants selecting the less likely option in cases where the reward value, which varied from trial to trial in their work, attached to that option was higher. Bland and Schaefer had a reward value which did not vary throughout the experiment, but participants often opted for the less likely outcome. One possible development of the computational models would have been to introduce a parameter controlling a rate of guessing as used by Steyvers *et al*. This might have given a more realistic approximation of human behaviour if an underlying Bayesian process was being used by the participants.

Behrens *et al* created an alternative Bayesian model which assumed that the reward rate, rather than varying continuously, would have a steady value and jump to any other steady value. They found that this model gave very similar performance to the one which assumed continuous changes in reward. Another possible alternative would have been to create a model which assumed that when a jump in reward occurred, the new reward would favour the opposite outcome to the old reward. This would more closely match the actual experiment.

Behrens *et al* suggest that "volatility is detected by subjects" and that "people both estimate and use this volatility parameter optimally, gauging the value of each new piece of information that they acquire." The analysis of the computational models has shown that volatility can be detected by a simpler model than that used by Behrens *et al*. There is no evidence from the data of Bland and Schaefer that participants optimally use an estimate of the volatility of the situation.

Bland and Schaefer considered the behaviour of participants on the trials immediately following those in which they had been informed that their response was incorrect, a loss. They found that participants were much more likely to switch behaviour after a loss than after a win. This was also the case with the Bayesian model which will never switch its behaviour after a win as it is always responding with the most likely outcome and a win only makes the current belief stronger. A simple strategy such as win-stay lose-shift will also result in more switches after a loss, but would not be an optimal strategy.

The Bayesian models implemented in this study required the storage of grids of probabilities. For the complex model, the processing of the data to generate responses took quite a long time. No actual timings were made, however. The simpler model which only had one hidden variable and a parameter was able to process the data more quickly than the complex model. The models also require much less computation if only two options are used for each random variable. This alternative was not compared with the performance of the models in which the discrete variables took 49 values.

In this study, the observed evidence was taken to be a single variable, equivalent to the underlying rule in the study by Bland and Schaefer. There were more possible variables which could have been added. One of these could have been the actual colour which was shown to the participants. At the beginning of the study, the participants would not know what information was relevant. It would have been a good enhancement of the Bayesian models produced to add additional variables, which in actual fact are irrelevant to the sequence and see if the model would start to ignore those variables.

One of the limitations of the psychological data used was that it was difficult to extract the learning of the underlying rule from the generation of a response. It would be interesting to introduce the experiment to the participants so that they were informed that the process is random and they are just required to indicate which is the most likely outcome. I would propose that the conditions of the experiment were kept as simple as possible to try to just extract information about how the participants were incorporating new information into existing beliefs.

The original aim of this study was to test the hypothesis that humans use Bayesian reasoning. Limitations in the human data available meant that it has not been possible to test this hypothesis. It has been demonstrated that human behaviour can be associated with two broad groups of behaviour, maximising and probability matching. It has also been shown that a reasonably simple Bayesian model performs as well as a more complex model in a learning task.

Bibliography

[1] A. Bland and A. Schaefer. How to maximize gains in uncertain contexts? electrophysiological evidence for dissociable modes of cognitive control in a challenging decision-making task. University of Leeds, 2010.

[2] T. E. J. Behrens, M. W. Woolrich, M. E. Walton, and M. F. S. Rushworth. Learning the value of information in an uncertain world. *Nature Neuroscience*, 10:1214 – 1221, 2007.

[3] A. Bland and A. Schaefer. Electrophysiological correlates of decision making under varying levels of uncertainty. University of Leeds, 2011.

[4] S. Siegel and D. A. Goldstein. Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology*, 57(1):37 – 42, 1959.

[5] B. R. Newell, D. A. Lagnado, and D. R. Shanks. *Straight choices : the psychology of decision making*. Psychology Press, Hove, 2007.

[6] D. R. Shanks, R. J. Tunney, and J. D. McCarthy. A re-examiniation of probability matching and rational choice. *Journal of Behavioral Decision Making*, 15:233 – 250, 2002.

[7] J. Baron. *Thinking and deciding*. Cambridge University Press, Cambridge, 4th edition, 2008.

[8] W. Gaissmaier and L. J. Schooler. The smart potential behind probability matching. *Cognition*, 109(3):416 – 422, 2008.

[9] D. J. Koehler and G. James. Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, 113(1):123 – 127, 2009.

[10] S.J. Russell and P. Norvig. *Artificial intelligence : a modern approach*. Pearson Education, 3rd edition, 2010.

[11] R.T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14:1–13, 1946.

[12] C.M. Bishop. *Pattern recognition and machine learning*. Springer, New York, 2006.

[13] D.L. Poole and A.K. Mackworth. *Artificial intelligence : foundations of computational agents*. Cambridge University Press, New York, 2010.

[14] F. V. Jensen. *An introduction to Bayesian networks*. UCL Press, London, 1996.

[15] T. L. Griffiths, C. Kemp, and J. B. Tenenbaum. Bayesian models of cognition. In Ron Sun, editor, *The Cambridge handbook of computational psychology*. Cambridge University Press, Cambridge, 2008.

[16] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[17] M. Steyvers, M. D. Lee, and E.-J. Wagenmakers. A bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, 53:168–179, 2009.

Appendix A

Personal Reflection

Using data collected by other people gave me an opportunity which would not have been available otherwise. Understanding human reasoning requires data from humans. This is often obtained through psychological studies which are time consuming and expensive to set up. Using a study which was designed by someone else did mean that I had no control over what was tested and how. The tests which had been carried out was probably more complicated than I would have preferred but still very valuable to me.

Studying two versions of a report submitted for publication by Bland and Schaefer [1, 3], gave me insight into how other people interpret and describe experimental data. Having the original data corresponding to a printed report is a rare privilege which allowed me to look at how conclusions were drawn from the data and to compare those to my own findings from the same data.

I found it more difficult than I had originally expected to replicate the work of Behrens *et al* [2]. There is a difference between reading a paper and understanding it well enough to reproduce computation. I needed to understand the paper well enough to make decisions as to how to build models where the details in a paper were not clear. This has also shown me the importance of detail and accuracy in reports.

Writing has taken me roughly 250 hours, which is a large proportion of a 600 hour project. This means that the experimental part of the project is actually small considering the elapsed time. A fellow student was told in his progress meeting that his assessor expects to take $1\frac{1}{2}$ days per page for academic writing. I felt heartened by this as it made me realise that I wasn't necessarily working too slowly. I carried out more experimentation than is included in this report. I had to decide what it was important to include.

I have often had difficulty writing my ideas in a precise way even when I had regularly made rough notes. I also found that when trying to describe my work, I had not always produced suitable graphs to illustrate the points that I wanted to make. Writing has helped me to focus on details and to organise my ideas.

I chose to use LaTeX to format my report due to the need to include mathematical equations. I had not previously used LaTeX and found that I needed quite some time getting used to the formatting techniques. I also found the text entry environment awkward to use, especially as the text could only be viewed in a very small format and there was no spell checker. In spite of these difficulties, I would still recommend other students to use LaTeX.

The level of planning which I used was deliberately not very detailed. The important factors were to not spend too long reading and actually start some practical work and to not spend so long carrying out experiments that there was no time to write up. I was glad that I allowed a block of time to work on my project full time during the Easter vacation. This gave me the chance to try out different ideas without having to worry about the assessment criteria at that stage. Within the broad framework I used for a project plan, I had a lot of freedom as to the actual work carried out. The only point at which I used a detailed plan was during the last two weeks of the project to make sure that appropriate time was spent on each section of the report.

In my interim report, I chose to cover my work up to that time with only a brief outline of what I expected to work on next. The comments from the assessor gave me the impression that all that he was looking for was details about what would be done next. I felt as if I had completely missed the purpose of the interim report although I had checked the project website to make sure that I had covered the required details. I would advise a future student in a similar situation to arrange their progress meeting as soon as possible. I am grateful to my assessor for the time he took over the progress meeting and I feel that I had explained my project to an appropriate level during that meeting.

In spite of some difficulties experienced trying to replicate a study, I would still recommend this as a starting point to others. I gained a lot of insight into Bayesian networks which allowed me to draw my own conclusions.

Appendix B

Materials used

Bland and Schaefer provided me with experimental data in a separate text file for each participant.

I programmed all the computational models myself.

Appendix C

Interim report

The interim report for this project is submitted with the hard copy only.

Appendix D

Project Plans

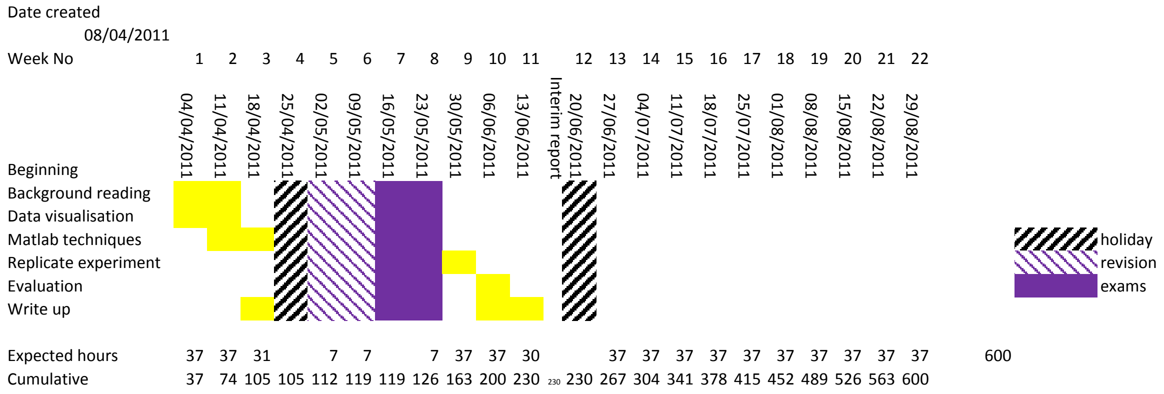
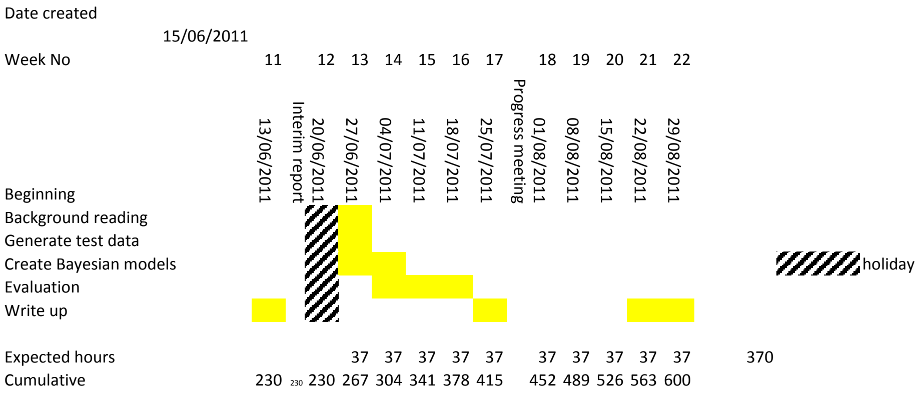
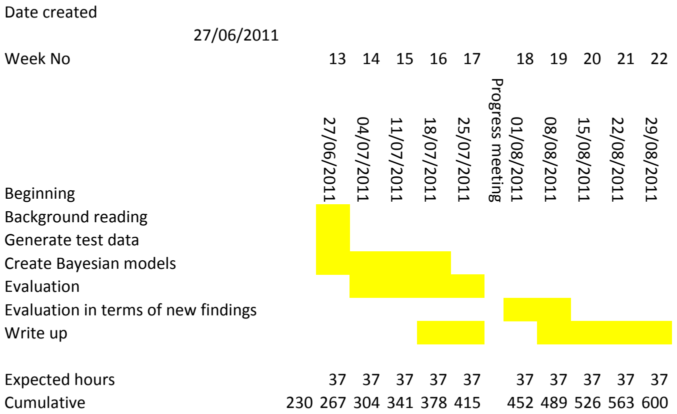


Figure D.1: Original plan



Date created																
15/06/2011																
Week No	11		12	13	14	15	16	17		18	19	20	21	22		
	13/06/2011	Interim report	20/06/2011	27/06/2011	04/07/2011	11/07/2011	18/07/2011	25/07/2011	Progress meeting	01/08/2011	08/08/2011	15/08/2011	22/08/2011	29/08/2011		
Beginning																
Background reading			▨	🟨												
Generate test data			▨	🟨												
Create Bayesian models			▨	🟨	🟨											▨ holiday
Evaluation			▨		🟨	🟨	🟨									
Write up	🟨		▨					🟨					🟨	🟨		
Expected hours				37	37	37	37	37		37	37	37	37	37	370	
Cumulative	230	230	230	267	304	341	378	415		452	489	526	563	600		

Figure D.2: Provisional plan at the time of submission of the interim report

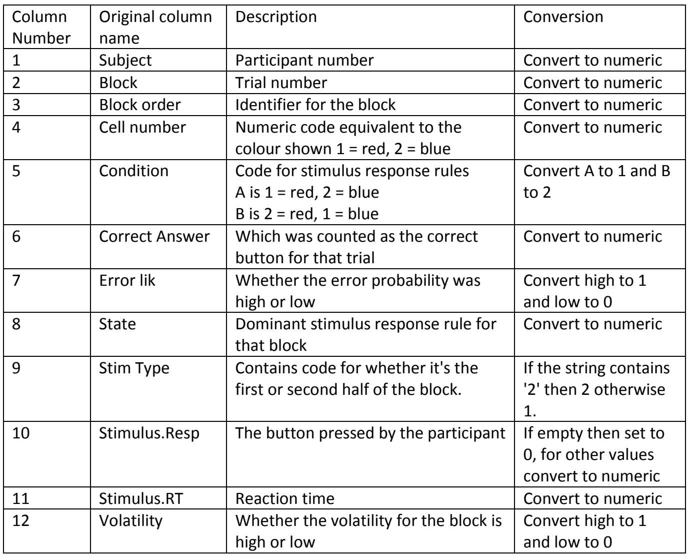


Date created												
27/06/2011												
Week No		13	14	15	16	17		18	19	20	21	22
		27/06/2011	04/07/2011	11/07/2011	18/07/2011	25/07/2011	Progress meeting	01/08/2011	08/08/2011	15/08/2011	22/08/2011	29/08/2011
Beginning												
Background reading		🟨										
Generate test data		🟨										
Create Bayesian models		🟨	🟨	🟨	🟨							
Evaluation			🟨	🟨	🟨	🟨						
Evaluation in terms of new findings								🟨	🟨			
Write up					🟨	🟨			🟨	🟨	🟨	🟨
Expected hours		37	37	37	37	37		37	37	37	37	37
Cumulative	230	267	304	341	378	415		452	489	526	563	600

Figure D.3: New plan

Appendix E

Data conversion



Column Number	Original column name	Description	Conversion
1	Subject	Participant number	Convert to numeric
2	Block	Trial number	Convert to numeric
3	Block order	Identifier for the block	Convert to numeric
4	Cell number	Numeric code equivalent to the colour shown 1 = red, 2 = blue	Convert to numeric
5	Condition	Code for stimulus response rules A is 1 = red, 2 = blue B is 2 = red, 1 = blue	Convert A to 1 and B to 2
6	Correct Answer	Which was counted as the correct button for that trial	Convert to numeric
7	Error lik	Whether the error probability was high or low	Convert high to 1 and low to 0
8	State	Dominant stimulus response rule for that block	Convert to numeric
9	Stim Type	Contains code for whether it's the first or second half of the block.	If the string contains '2' then 2 otherwise 1.
10	Stimulus.Resp	The button pressed by the participant	If empty then set to 0, for other values convert to numeric
11	Stimulus.RT	Reaction time	Convert to numeric
12	Volatility	Whether the volatility for the block is high or low	Convert high to 1 and low to 0

Figure E.1: Columns selected and conversion

Appendix F

Manual Validation

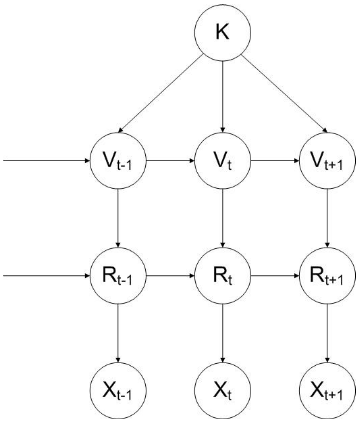


Figure F.1: HMM variation with two hidden variables and one parameter

The graphical representation for the complex model is repeated above, Figure F.1. The equation for inference for this model is

$$\mathbf{P}(R_{t+1},V_{t+1},K|x_1,....,x_{t+1}) =$$
$$\alpha\mathbf{P}(x_{t+1}|R_{t+1})\sum_i(\mathbf{P}(R_{t+1}|R_t = r_i,V_{t+1})\sum_j\mathbf{P}(V_{t+1}|V_t = v_j,K)\mathbf{P}(R_t,V_t = v_j,K|x_1,....,x_t))$$

Considering each variable to have just two values, low and high, represented as *l* and *h* respectively, the transition matrices can be defined as follows

Transition from R_t to R_{t+1} for each value of V_{t+1}

$$V_{t+1} = l \qquad\qquad V_{t+1} = h$$

$$\begin{array}{c|cc} & R_t = l & R_t = h \\ R_{t+1} = l & 0.75 & 0.25 \\ R_{t+1} = h & 0.25 & 0.75 \end{array} \qquad \begin{array}{c|cc} & R_t = l & R_t = h \\ R_{t+1} = l & 0.9 & 0.1 \\ R_{t+1} = h & 0.1 & 0.9 \end{array}$$

The transition from V_t to V_{t+1} is given as follows

$$K = low \qquad\qquad K = high$$

$$\begin{array}{c|cc} V_{t+1} \diagdown V_t & low & high \\ low & 0.9 & 0.1 \\ high & 0.1 & 0.9 \end{array} \qquad \begin{array}{c|cc} V_{t+1} \diagdown V_t & low & high \\ low & 0.6 & 0.4 \\ high & 0.4 & 0.6 \end{array}$$

The output is determined from the hidden variable as follows

$$\begin{array}{c|cc} & X = 1 & X = 2 \\ R_t = l & 0.8 & 0.2 \\ R_t = h & 0.2 & 0.8 \end{array}$$

The prior probability distribution is taken to be uniform.

First Iteration

Time $t = 1$

The first summation from the right updates the probability distribution for volatility based on its transition matrix. For the first step, this is $\mathbf{P}(R_0, V_1, K) = \sum_j \mathbf{P}(V_1 | V_0 = v_j, K)\mathbf{P}(R_0, V_0 = v_j, K)$. The calculations are shown separately for each value of K.

$$K = l \quad \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \begin{pmatrix} 0.125 & 0.125 \\ 0.125 & 0.125 \end{pmatrix} = \begin{array}{c|cc} & R_0 = l & R_0 = h \\ V_1 = l & 0.125 & 0.125 \\ V_1 = h & 0.125 & 0.125 \end{array}$$

$$K = h \quad \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 0.125 & 0.125 \\ 0.125 & 0.125 \end{pmatrix} = \begin{array}{c|cc} & R_0 = l & R_0 = h \\ V_1 = l & 0.125 & 0.125 \\ V_1 = h & 0.125 & 0.125 \end{array}$$

Now, compute the second sum using the matrices determined above. $\sum_i (\mathbf{P}(R_1 | R_0 = r_i, V_1)\mathbf{P}(R_0, V_1, K)$

$$V_1 = l \qquad \begin{pmatrix}0.9 & 0.1\\ 0.1 & 0.9\end{pmatrix}\begin{pmatrix}0.125 & 0.125\\ 0.125 & 0.125\end{pmatrix} = \begin{array}{ccc} & K=l & K=h \\ R_1 = l & 0.125 & 0.125 \\ R_1 = h & 0.125 & 0.125 \end{array}$$

$$V_1 = h \qquad \begin{pmatrix}0.75 & 0.25\\ 0.25 & 0.75\end{pmatrix}\begin{pmatrix}0.125 & 0.125\\ 0.125 & 0.125\end{pmatrix} = \begin{array}{ccc} & K=l & K=h \\ R_1 = l & 0.125 & 0.125 \\ R_t = h & 0.125 & 0.125 \end{array}$$

At this point a prediction is made. As no observations have yet been made, the joint probability distribution between R, K and V is still uniform. The default prediction of '1' is made.

A '1' is observed so each of these is multiplied by 0.8 for probabilities corresponding to $R = l$ and by 0.2 for $R = h$.

$$V = l \qquad \begin{pmatrix}0.125 & 0.125\\ 0.125 & 0.125\end{pmatrix} \times \begin{pmatrix}0.8 & 0.8\\ 0.2 & 0.2\end{pmatrix} = \begin{pmatrix}0.1 & 0.1\\ 0.025 & 0.025\end{pmatrix}$$

$$V = h \qquad \begin{pmatrix}0.125 & 0.125\\ 0.125 & 0.125\end{pmatrix} \times \begin{pmatrix}0.8 & 0.8\\ 0.2 & 0.2\end{pmatrix} = \begin{pmatrix}0.1 & 0.1\\ 0.025 & 0.025\end{pmatrix}$$

The resulting matrices now have to be normalised so that the probabilities again add to one. After normalisation, the joint probability distribution which is carried to the next iteration is as follows.

$$V_1 = l \qquad\qquad\qquad\qquad V_1 = h$$

$$\begin{array}{ccc} & K=l & K=h \\ R_1 = l & 0.2 & 0.2 \\ R_1 = h & 0.05 & 0.05 \end{array} \qquad\qquad \begin{array}{ccc} & K=l & K=h \\ R_1 = l & 0.2 & 0.2 \\ R_1 = h & 0.05 & 0.05 \end{array}$$

Second Iteration

Updating with the transition for V, the first summation becomes $\mathbf{P}(R_1, V_2, K) = \sum_j \mathbf{P}(V_2|V_1 = v_j, K)\mathbf{P}(R_1, V_1 = v_j, K|x_1)$.

The matrices above represent a 3D grid. This grid has to be oriented correctly so that the matrix multiplication gives summation over V_1. Rotating grids about different axes is available in Matlab.

$$K = l \qquad \begin{pmatrix}0.9 & 0.1\\ 0.1 & 0.9\end{pmatrix}\begin{pmatrix}0.2 & 0.05\\ 0.2 & 0.05\end{pmatrix} = \begin{array}{ccc} & R_1=l & R_1=h \\ V_2 = l & 0.2 & 0.05 \\ V_2 = h & 0.2 & 0.05 \end{array}$$

$$K = h \qquad \begin{pmatrix}0.6 & 0.4\\ 0.4 & 0.6\end{pmatrix}\begin{pmatrix}0.2 & 0.05\\ 0.2 & 0.05\end{pmatrix} = \begin{array}{cc} & \begin{array}{cc} R_1 = l & R_1 = h\end{array}\\ \begin{array}{c} V_2 = l\\ V_2 = h\end{array} & \begin{pmatrix}0.2 & 0.05\\ 0.2 & 0.05\end{pmatrix}\end{array}$$

These are then used in the second summation, updating with the transition for R

$$V_2 = l \qquad \begin{pmatrix}0.9 & 0.1\\ 0.1 & 0.9\end{pmatrix}\begin{pmatrix}0.2 & 0.2\\ 0.05 & 0.05\end{pmatrix} = \begin{array}{cc} & \begin{array}{cc} K = l & K = h\end{array}\\ \begin{array}{c} R_2 = l\\ R_2 = h\end{array} & \begin{pmatrix}0.185 & 0.185\\ 0.065 & 0.065\end{pmatrix}\end{array}$$

$$V_2 = h \qquad \begin{pmatrix}0.75 & 0.25\\ 0.25 & 0.75\end{pmatrix}\begin{pmatrix}0.2 & 0.2\\ 0.05 & 0.05\end{pmatrix} = \begin{array}{cc} & \begin{array}{cc} K = l & K = h\end{array}\\ \begin{array}{c} R_2 = l\\ R_2 = h\end{array} & \begin{pmatrix}0.17 & 0.17\\ 0.08 & 0.08\end{pmatrix}\end{array}$$

At this point a prediction can be made. This is done by summing over V and K to give a probability distribution over R. This gives $\mathbf{P}(R_2) = (0.710.29)$. A prediction of ‘1’ is made.

A ‘1’ is observed.

$$V_2 = l \qquad \begin{pmatrix}0.185 & 0.185\\ 0.065 & 0.065\end{pmatrix} \times \begin{pmatrix}0.8 & 0.8\\ 0.2 & 0.2\end{pmatrix} = \begin{pmatrix}0.148 & 0.148\\ 0.013 & 0.013\end{pmatrix}$$

$$V_2 = h \qquad \begin{pmatrix}0.17 & 0.17\\ 0.08 & 0.08\end{pmatrix} \times \begin{pmatrix}0.8 & 0.8\\ 0.2 & 0.2\end{pmatrix} = \begin{pmatrix}0.136 & 0.136\\ 0.016 & 0.016\end{pmatrix}$$

After normalisation, this gives

$$\begin{array}{cc} V_2 = l & V_2 = h \\ \\ \begin{array}{cc} & \begin{array}{cc} K = l & K = h\end{array}\\ \begin{array}{c} R_2 = l\\ R_2 = h\end{array} & \begin{pmatrix}0.236 & 0.236\\ 0.021 & 0.021\end{pmatrix}\end{array} & \begin{array}{cc} & \begin{array}{cc} K = l & K = h\end{array}\\ \begin{array}{c} R_2 = l\\ R_2 = h\end{array} & \begin{pmatrix}0.217 & 0.217\\ 0.026 & 0.026\end{pmatrix}\end{array}\end{array}$$

Third Iteration

$$K = l \qquad \begin{pmatrix}0.9 & 0.1\\ 0.1 & 0.9\end{pmatrix}\begin{pmatrix}0.236 & 0.021\\ 0.217 & 0.026\end{pmatrix} = \begin{array}{cc} & \begin{array}{cc} R_2 = l & R_2 = h\end{array}\\ \begin{array}{c} V_3 = l\\ V_3 = h\end{array} & \begin{pmatrix}0.234 & 0.022\\ 0.219 & 0.025\end{pmatrix}\end{array}$$

$$K = h \quad \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 0.236 & 0.021 \\ 0.217 & 0.026 \end{pmatrix} = \begin{array}{ccc} & R = l & R = h \\ V_3 = l & 0.228 & 0.023 \\ V_3 = h & 0.225 & 0.024 \end{array}$$

$$V_3 = l \quad \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \begin{pmatrix} 0.234 & 0.228 \\ 0.022 & 0.023 \end{pmatrix} = \begin{array}{ccc} & K = l & K = h \\ R_{t+1} = l & 0.213 & 0.208 \\ R_{t+1} = h & 0.043 & 0.043 \end{array}$$

$$V_3 = h \quad \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix} \begin{pmatrix} 0.219 & 0.225 \\ 0.025 & 0.024 \end{pmatrix} = \begin{array}{ccc} & K = l & K = h \\ R_3 = l & 0.171 & 0.175 \\ R_3 = h & 0.074 & 0.074 \end{array}$$

A ‘2’ is observed

$$V = l \quad \begin{pmatrix} 0.213 & 0.208 \\ 0.043 & 0.043 \end{pmatrix} \times \begin{pmatrix} 0.2 & 0.2 \\ 0.8 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.043 & 0.042 \\ 0.034 & 0.034 \end{pmatrix}$$

$$V = h \quad \begin{pmatrix} 0.219 & 0.225 \\ 0.025 & 0.024 \end{pmatrix} \times \begin{pmatrix} 0.8 & 0.8 \\ 0.2 & 0.2 \end{pmatrix} = \begin{pmatrix} 0.034 & 0.035 \\ 0.059 & 0.059 \end{pmatrix}$$

Normalisation gives,

$$V_3 = l \qquad\qquad\qquad\qquad V_3 = h$$

$$\begin{array}{ccc} & K = l & K = h \\ R_3 = l & 0.126 & 0.124 \\ R_3 = h & 0.1 & 0.1 \end{array} \qquad\qquad \begin{array}{ccc} & K = l & K = h \\ R_3 = l & 0.1 & 0.103 \\ R_3 = h & 0.174 & 0.174 \end{array}$$