The effects of goals on forecasting error in chaotic situations using a control task

Lau, Kam Keung

The Effects of Goals on Forecasting Error in Chaotic Situations using a control task

Kam Keung LAU

Submitted for the Qualification of

Master of Arts

University of Durham

Department of Psychology

September 1998

ABSTRACT

Two experiment were designed to investigate the cognitive patterns that people used in learning to forecast, specifically, the effects on forecasting produced by different learning goals. A control task formerly employed by Harvey (1990) was used in the present experiment. Participants were instructed to interact with computers and engaged in a mock medical decision-making task, in which patients with mood disorders were treated by psychiatrists. The pattern of mood disorders was characterised by a chaotic formula. Participants were divided into groups having different goals. The results showed that goals has a significant effect on learning, which affected the performance of participants when making judgements and the patterns in forecasting. The results were discussed in terms of the learning types of either rule searching or problem solving. With a problem solving goal, participants tended to learn instances; with a rule searching goal, participants tended to learn rules.

ACKNOWLEDGEMENT

I would like to thank Dr. Rosemary Stevenson for her advice and patience during the period of supervision and Bob Metcalf for his effect in helping us to work out the computer programs of the experiments.

TABLE OF CONTENTS

DECLARATION AND STATEMENT OF COPYRIGHT

Declaration

The research contained in this thesis was carried out by the author between October 1997 and September 1998 whilst a postgraduate in the Department of Psychology at the University of Durham. None of the work contained in this thesis has been submitted in candidature for any other degree.

Statement of Copyright

Chapter One

Introduction

INTRODUCTION

Forecasting and everyday life

Most decisions concern the future and these decisions will usually constitute some consequences for the future. People have to consider different consequences of their decisions and actions for the future. This kind of behaviour is called forecasting. Since no one can foresee the future, forecasting involves making decisions in the presence of uncertainty. For the same reason, forecasting involves both reasoning and judgement: reasoning about the relationship between action and consequence and judgement by considering subjective values and utilities. People have to make numerous forecasts every day. When one gets up in the morning, one has to forecast the transport situation so as to select the best means of travel. In addition, one has to forecast the weather, so that one can decide what to wear. When a person enters the workplace, he/she has to make forecasts about the reactions of other people so that he/she knows how to behave toward others.

Good forecasting will bring abundant benefits to our lives. With good forecasts, people can utilise their money, time and effort properly and get the desirable outcome. People can determine the best way to invest their money and make profit. They learn how to manage their time so that they will not waste their effort. On the other hand, poor forecasting will greatly diminish the quality of lives. Too much effort is used in futile ways. Money is spent in lost situations. Time is used without

considering the real situation in the future and so wasted. In modern society, different classes of people have to make different types of forecasting. In business, forecasting plays the role of making profitable budget and cost-price estimations. Within a government, policy makers have to do accurate forecasting about the impact and effect of various policies so that the social resources will not be wasted. Within economics, forecasting is done by various authorities in order to make financial plans. Within interpersonal relationships, people usually have to make some forecasts about other people's reactions in everyday life.

Human activities are of various types. For some activities, the behaviour can be described by rules (e.g. using computer programs) while others cannot (e.g. forecasting the weather). The patterns of some activities are obvious while some are not. For those activities with an implicit pattern, people cannot predict the exact behaviour. (For example, one may know the personality of a friend to certain extent but cannot forecast what the friend's mood will be tomorrow.) However, they may know the pattern to certain extent. Therefore, to some extent, knowing the pattern behind an activity enables people to make predictions about the future, and thus a grasp of the pattern is an indispensable factor in forecasting. The basic approach of this kind of forecasting is observing the past as a guide of future (Phillips 1987). For example, politicians forecast the trend of public issues by observing similar previous cases. In that sense, the ability to forecast is connected with learning. In the present study, the relationship between learning and forecasting is examined.

Research on general forecasting

Most modern research on forecasting emphasises one of two lines of investigation. One line is to work out a statistical model to make a rational and accurate model of forecasting. It involves various types of mathematical calculation to examine the trend of a phenomenon and make predictions about the future. Another line of investigation is the psychology of decision making in cognitive psychology. It concerns subjective utility and the values of the person making the judgement. It makes comparisons between the subjective perceived value and the objective expected value of the outcome. For the statistical approach to forecasting, a number of forecasting methods have been devised. One important method is the time-series analysis which is a method for describing, observing and predicting the trend of a phenomenon. Basically, its logic is to collect quantitative information about a phenomenon throughout a period of time up to the time of forecasting. Then by observing the data in the graph the prediction is made. Another important statistical approach is the Bayesian prediction method. It makes a complete model of a set of observable values and then calculates the probability distribution of the unobserved values conditional on those already observed (Geisser 1993).

In cognitive psychology, the researchers mainly investigate the interaction between objective expected values and subjective utility as well as irrational biases. The focus of investigation is the inconsistency between the subjective perception of some outcome value and the actual expected values obtained by considering the probability of the outcomes. The research shows that human decision making consists

of heuristics and various kinds of biases in judgement. One of them is the belief bias which shows that people fail to follow deductive logic because they are influenced by prior knowledge. Another important bias is the confirmation bias which shows that in the presence of uncertainty people are motivated to confirm their own belief and so fail to make rational decisions (Evans 1987). Both forecasting models and forecasting psychology are good resources for studying forecasting. The first approach is the rational model of forecasting and provides a normative view of forecasting (the ideal way of forecasting). The latter approach is human decision making and judgement and provides a descriptive view of human forecasting (the actual way of forecasting). However, little concern is given to a prescriptive model of forecasting: the study of how to improve forecasting and by which method, as well as the factors affecting improvement.

In the present decade, a new method of forecasting, the method of chaos theory, has been developed in social science. It has successfully explained various phenomena in the natural sciences and subsequently been used to develop some new paradigms of science. It is part of a relatively new and rapidly expanding area of science which demonstrates how apparently random phenomena can stem from a simple set of deterministic laws. Social sciences, historically, have followed both the intellectual and methodological paradigms of the natural sciences. The uses of chaos theory are the latest effort by social scientists to combine and apply the theory and method from natural sciences. Most importantly, chaos theory appears to provide a means for understanding and examining many of the uncertainties and unpredictable aspects of certain social systems and behaviours. It provides a valuable reference for

the investigation of human behaviour. Existing examinations of chaos theory can be found in economics, political science, psychology etc.. Within psychology, the concept of chaos theory is beginning to be applied to different areas by researchers from cognitive, developmental and clinical aspects (Ayers 1997).

The scope of the present study

Existing statistical analyses of forecasting have not considered the human factor in forecasting. Although they provide models for forecasting, most of the raw forecasts from these models are subjected to judgmental revision. Various studies have shown that the revised forecasts are better than the raw ones (Harvey, Bolger & McClelland 1994). Thus, it is important to know the processes underlying the psychology of forecasting. On the other hand, existing decision making theories in cognitive psychology have not been extended to the study of forecasting. Existing theories of decision making mainly examine present events. Although, occasionally, some future events are mentioned, the events have a known probability. Forecasting involves future events in which the probability of the outcome is uncertain. A study investigating the psychology of forecasting has examined the underlying cognitive processes and provided some understandings in that area (Harvey, Bolger & McClelland 1994). In the experiment, a series of criminal rate records in an underground railway system was employed. Participants were given successive data points of the criminal rate and were asked to predict the next figures. It was found that participants made their predictions basing on the trend of the previous data. However, the experiment did not involve any

manipulation of factors which is the interest of the present research.

There are two research foci in the present study. One is on the psychology of forecasting and the factors affecting forecasting behaviour. Recent studies using control tasks on the effect of goals on learning provide some insights for the present study. The effect of goals will be examined to see how it influences forecasting. Control tasks will be used as tools to make the investigation. The advantage of using a control task is that it produces continuous and meaningful quantitative data for investigation. In addition, people confront situations that are similar to control systems all around the real world (in biology, economics, engineering and interpersonal relationships, etc.).

Another research focus rests on the psychology of forecasting under chaotic phenomena. Since chaos theory is a relatively new concept in cognitive psychology, related areas have not been thoroughly and systematically explored. Existing research into the application of chaos theory in psychology focuses mainly on the clinical areas of neuroscience, psychophysiology and psychopathology, and mainly concerns the relation between neuroscience and cognition. Research using chaos theories that has investigated human thinking and decision making has been extremely limited, and the findings can be plausibly explained by simple heuristics or the use of associative memories. The studies have been criticised for not being concerned with the nature of chaos phenomena (Ayers 1997). However, since chaos theory has been successfully used to explain not only some scientific phenomena but also some important and applicable social phenomena, such as economics in modern societies, it is worth

exploring how people respond and forecast in chaotic situations. Such studies may provide relevant policy makers with a valuable means of better understanding the implications of policies on chaotic social systems.

The research on control tasks

Control mechanisms are widespread in nature. Living organisms apply the mechanisms to maintain essential variables like body temperature and blood sugar levels at desired points. Today, people use the same concepts to develop various engineering systems in the world (Sontag 1990). Also, governments have to control interest rates and taxes to keep inflation and unemployment within acceptable ranges. Businessmen have to control the prices, wages and stock so as to maximise their profit (Harvey 1990). A control task is a system to which people give a certain value of input and then the system will produce a corresponding value of output. Normally, there should be some relationship between the system input and system output so that the relationship can be represented in the form of a formula in which stimulus and response form a special relationship. (For example, in many well-known situations, the stronger a stimulus, the greater the response.) There are different types of control systems. A control system is said to be either dynamic or static. A dynamic control system will have some cumulative effect of the previous input, while a static control system will produce outputs that are independent of previous inputs. Another kind of category is based on the stability of the system. A stable control system is a deterministic system without noise while an unstable system will be affected by the presence of noise. Noise will

alter the output of a system so that the system will not have stable output values.

The studies reviewed below all used dynamic control systems. Harvey and his colleagues examined quantitative forecasting; the main focus of investigation by Berry and Broadbent was on the characteristics of the learning process, while Geddes and Stevenson (1997) examined the effects of goal on learning a dynamic control task.

Harvey's studies of control tasks

Harvey's work provides an informative background to the present investigation. In a series of studies (Harvey 1990; Harvey, Koehler & Ayton 1997; Koehler & Harvey 1997), logistic map control systems were given to the participants. The control systems were governed by the following formula with or without noise:

$$Y_{t+1} = A\, Y_t\, (\, 1 - Y_t\,)$$

where Y and A were the output and parameter of the system respectively and Y_t was the previous output while Y_{t+1} was the present output of the system. The systems were discrete versions of the well-known growth curve equation. The logistic map had several important characteristics, its ubiquity, simplicity and stability and, interestingly, different behaviours that it produces for different values of parameter A. Considering the noise-free version of the system, when A is greater than 1.0 but less than or equal to 3.0, Y has a single asymptotically stable value that increases with A. Between 3.0

and 3.57 of A, Y would alter between two or more values. For further increases in A, the system output of Y would be chaotically unpredictable.

In addition to the principle behind the system, a problem situation was presented to the participants, which was a doctor-drug-dosage problem. Participants were told that they had to imagine that they were psychiatrists specialising in the treatment of people with mood disorders. For diagnostic purposes, their patients keep diaries of the number of experiences that make them feel noticeably happy. For normal people, it is known that this happiness index was between 29 and 31 per month. Patients' moods would be outside this range initially but participants were to bring them into it by providing appropriate treatment. Depressive patients must be treated with a suitable amount of antidepressant drug. Manic patients are more irresponsible and boisterous than normal people while manic-depressives' moods oscillate between the extremely depressed and extremely boisterous in a predictable manner. Both manic and manic-depressive patients must be treated with the drug lithium. In treating patients, a psychiatrist must try to find out the most appropriate dosage for bringing each patient's moods into the normal range.

In Harvey's (1990) experiment, the stimuli were the different mood disorder situations. The mock patient cases were generated by computer using the above logistic map formula of the control system. The Y values of the formula acted as the happiness indexes of the patients while the parameter A values indicated the severity and pattern of a patient's illness. A random number between zero and one was chosen as the first value of Y while parameter A was randomly selected from a list of eight

numbers (2.0, 2.2, 2.9, 3.1, 3.3, 3.5, 3.7 and 3.9). The system was then iterated 100 times to clear any transient responses. The next two iterations generated were the initial data of a patient's happiness index. This information was shown on a computer screen. For a patient having a happiness index of 24 and 40 for the first two months, the information 'MONTH 1=24' would be shown under the heading 'HAPPINESS INDEX' and 'NONE' under the heading 'TREATMENT REGIME' as the first row of data on the screen. Similarly, the information 'MONTH 2 = 40' under the heading 'HAPPINESS INDEX' and 'NONE' under the heading 'TREATMENT REGIME' would be shown on the second row. Participants sat in a room facing the computer screens and had to decide which drug and dosage level should be given to the patient. Each participant was asked to make their response by using a joystick and pressing a button on its handset. By moving the joystick, all possible answers to the current question appeared on the screen to the right of the question. They were asked to move the joystick until the answer that they required was presented and then pressed a button on the handset. This answer was recorded by the computer and the next question then appeared on the screen.

A major line of inquiry behind Harvey's studies is to understand the pattern of behaviour in judgmental forecasting with the control task. One focus of Harvey's work (Harvey 1990) was a descriptive study of judgmental control behaviour in which the ability and pattern of controlling were investigated. The major finding was that people have the ability to grasp the pattern of certain cases of the control task. However, the ability to control was highly dependent on the complexity of the tasks. In relatively unpredictable (chaotic) cases, participants did worse; while in relatively stable cases,

participants did better.

Another focus was the difference in forecasting between executors and observers. In a series of experiments (Harvey, Koehler & Ayton 1997), overconfidence of decision effectiveness was compared between executors and observers by using the medical decision task. Participants were divided into two groups. One group played the role of doctor (executors) while another group acted as nurse (observers). Each participant in the nurse group was matched with one participant in the doctor group. Information about the control tasks was given to both groups.

In the first experiment, the doctor group participants made decisions about treatment according to the patients' mood disorder. The nurse group participants were then informed of the decisions. Both groups had to evaluate the adaptability of the decisions. It was found that in general both groups tended to be overconfident in the sense of overestimating the probability that the treatment would be effective. However, the doctor group participants were also significantly more overconfident than the nurse group participants. In a series of further experiments, it was found that the critical variable involved in the difference between executors' and observers' overconfidence was whether or not feedback on the decision was received. A group showed significantly greater overconfidence in its own decision whenever that group received feedback. It was also found that, in this series of experiments, there was no significant difference in treatment efficacy between the prescription judgements of those executor and observer groups.

Berry & Broadbent's dynamic control task

Several studies on control tasks have been done by Berry and Broadbent. The most related one is the series of studies in 1984. Three experiments explored the relationship between performance on a control task and the explicit or reportable knowledge associated with that performance. The researchers examined how this relationship was affected by task experience, verbal instruction and concurrent verbalisation. It was shown that practice could significantly improve the ability of controlling semi-complex computer-implemented systems but it had no effect on the ability to answer related questions. In contrast, verbal instruction significantly improved the ability to answer questions but had no effect on control performance. Verbal instruction combined with concurrent verbalisation led to a significant improvement in control scores. Verbalisation alone, however, had no effect on either task performance or question answering.

In one version of the experiment, participants were instructed to imagine that they were in charge of a sugar production factory in an underdeveloped country. They were told that, by ignoring all other factors, they could control the rate of production simply by changing the size of the work force. The size of the work force could be varied in twelve discrete steps (100, 200, 300, ... and 1200). The level of production and the work force had a quantitative relation, which was described by the equation, P = 2 x W-Pl, where W was the number between one and twelve representing the size of the work force, P1 was the number, again between one and twelve, representing the previous sugar output and P was the number representing the current sugar output. In

another version of the experiments, participants were told that they would be meeting a computer person named Clegg. 'They would communicate with Clegg through a typewriter keyboard and VDU. There were twelve possible grades of behaviour, these being Very Rude, Rude, Very Cool, Cool, indifferent, Polite, Very Polite, Friendly, Very Friendly, Affectionate, Very Affectionate and Loving. These adjectives had numerical equivalents, which corresponded to the twelve levels of work force and sugar output. Clegg initiated the interaction by displaying one of the three adjectives centred on Polite. Participants were instructed to shift Clegg's behaviour to the Very friendly level and to maintain it at that level. On each trial Clegg's and the participant's responses were displayed on VDU. These scrolled up the screen so that on any one trial it was possible to see the responses made on the previous six trials.

Participants were then asked both multiple-choice and open-ended questions. The former included prediction questions, which asked the possible response of the system given previous values of the input. For the open-ended question, participants were asked to describe how they had reached and maintained their target values. They were encouraged to write freely and not to worry about wording or grammaticality. The most important finding related to the present study is that these experiments tell us about the hypothesised "performance/verbalizable knowledge" dissociation. It was found that practice significantly improved the ability to control the sugar production and personal interaction tasks, but it did not have any effect on the ability to answer related questions. In contrast, verbal instruction significantly improved the ability to answer questions, yet it had no effect on control performance. One interpretation of these findings was that these tasks under certain conditions can be learned in some

implicit manner with no verbal awareness on the part of learners. However, is there a more fundamental element affecting learning through practice? What affects whether learning will be implicit and unreportable or explicit and reportable? One recent study has suggested that the choice of learning goal is one of the important factors (Geddes and Stevenson 1997).

Effect of learning goal

A goal is a state a person wants to achieve. In different aspects of life, there are different types of goals. In sport, athletes set proper goals as reference points for practising. In problem solving, problem solvers first have to identify their goals. The effects yielded by a goal are varied. In some cases, the goal produces an indispensable effect on achievement. The goals of athletes help them to decide whether to try harder or just to maintain the current performance. The goals of problem solvers help them to avoid wasting of effort, time and resource. Concerning learning, hitherto, there has been no systematic research on the effect of goal. For the present discussion, we look at some learning theories. There are various learning theories in the literature including those in language learning, conditioning and thinking. However, the most relevant learning theories for the control tasks are the theories concerning the process of acquiring knowledge of perceived uncertain systems. In the following text, relevant theories of learning will be introduced: rule learning/ instance learning and implicit learning/ explicit learning.

In some of the literature on of learning, researchers propose that there exist two types of learning: implicit learning and explicit learning. The term "implicit learning" was defined as the acquisition of complex abstract knowledge that takes place without the learner's awareness that he or she is learning (Berry and Broadbent, 1984; Hayes and Broadbent, 1988; Reber, 1967). Knowledge acquired during implicit learning cannot be reported explicitly. The learners are not aware of either the specific learning experiences or that they have retrieved the correct information. In contrast, explicit learning proceeds with human awareness, and the knowledge that is acquired is verbally reportable. Some researchers have suggested that implicit and explicit learning can be functionally dissociated (e.g. Berry and Broadbent, 1984; Hayes and Broadbent, 1988; Reber, 1967). However, other researchers have implied that the two learning systems are used in combination (Buchner, Funke and Berry, 1995; Dienes and Fahey, 1995; Geddes and Stevenson, 1997).

Another perspective on the learning of control tasks is the distinction between rule learning and instance learning. The inter-relationship between these two types of learning was investigated by Geddes and Stevenson. They showed that learning goals affect the extent to which each kind of learning predominates. As the name implies, rule learning means learning the pattern of the control system by examining and grasping the rule and underlying principle and instance learning means learning the task by memorising the instances that have appeared to the person in the learning trials. The rule-based process usually requires more deliberate or mental work to learn and make decisions while instance-based processes use mainly memory. The distinction between rule-learning and instance learning has been modelled as a dual space in the area of

hypothesis testing. Simon and Lea (1974) proposed that the problem space is separated into two conceptual spaces: a rule space and an instance space. When trying to reach a specific goal in problem solving, instance space is searched. However when testing hypotheses, both rule space and instance space are searched. That means explicit rules (or hypotheses) are first generated in rule space, and these rules are then tested by 'experiments' that generate states in instance space. Thus, in the dual space theory, problem solving takes place in instance space while hypothesis generation and testing takes place in both spaces.

The ideas of implicit/explicit learning and rule/instance learning seem to be closely related. One recent study has suggested that the effect of goal is one of the important factors to affect the process of learning (Geddes and Stevenson, 1997). In this study, Geddes and Stevenson proposed that a specific learning goal encourages instance learning while a non-specific learning goal encourages explicit hypothesis generation and testing. The two types of learning are referred to as instance learning and rule learning rather than implicit and explicit.

Geddes and Stevenson examined the effect of goals on learning processes and the major findings were the different effects produced by the goal. The same task of computer-person interaction (Clegg) as in Berry & Broadbent's 1984 experiment was given to the participants. Some of the participants were given the specific goal to maintain the system at a certain level and the other participants were given a non-specific goal to discover the pattern underlying Clegg's behaviour. It was found that the non-specific group were more able to grasp the rule behind the system. They could

report it explicitly and they could predict Clegg's next response when presented with a sequence of three responses, regardless of whether or not they had seen the sequences before. By contrast, the specific goal group seemed to learn instances. They could not describe the rule and they could only make correct predictions from sequences of three responses that they had seen before, they could not make correct prediction from novel sequences. Since the prediction questions can be seen as a test of forecasting, these latter results provide additional background information about the effects of learning on the ability to forecast. Thus, the distinction between rule and instance learning is important for knowing how participants will behave while learning to forecast.

Controlling and observing

In some control task studies, participants were requested to observe the controlling process of particular participants instead of controlling the systems. One reason for investigating the difference between observers and executors is that such studies have important implications for management research. It is important to ensure that the different groups of participants are exposed to the same amount of information so that the differences between groups can only be accounted for by the effects which were the focus of investigation not by the differences in knowledge.

For example, the specific goal group in Geddes' experiment tended to produce a restricted range of responses when controlling the task. By contrast, the non-specific goal group used a wide range of responses when searching for the underlying pattern.

Geddes ruled out the difference in range of response as a possible reason for the learning differences by showing that the same result obtained when observers with either a specific or non-specific goal saw the responses of a model who interacted with the computer and who also had either a specific or a non-specific goal. Only the goal of the observer had any significant effect on the learning outcomes. The observer design was also used in the present study to prevent the possible differences in knowledge between comparison groups.

As discussed in the findings of Harvey et al's series of experiments, (Harvey, Koehler & Ayton 1997) there were no significant differences in the ability to make judgements between executors and observers. The major difference between executors and observers was the degree of over-confidence. Executors were found to have a higher tendency to be over-confident in the absence of alternative recommendations. However, in the presence of feedback for the observers, the differences between executors and observers were eliminated. Thus, as was the case in Geddes' research, observation and execution of a control task produced comparable results.

Summary of review

Geddes's major finding of the effect of goal on the Clegg control task is that specific goal participants learned instances while non-specific goal participants learned rules. However, will the same effects happen in other control tasks? In the Clegg control task examined by Geddes, the computer provides twelve distinct states of behaviour to

imitate the mood of human beings. It is possible that in the Clegg experiment, participants were not given any motivation to learn and so the goal effects may be quite different when a practical problem is presented to the participants. One possibility is that a practical problem may motivate participants to investigate the rule behind the system so that they will learn the rules, regardless of their learning goal. Another possibility is that the practical problems will consolidate the goal effect and make the difference between different groups more manifest. Another point is that Clegg produced distinct qualitative states, will the same result obtain in a control task of quantitative forecasting? That is, will the qualitative findings in the control task also apply to quantitative tasks?

For the control task studies, Berry and Broadbent found that practice has a significant effect on the ability to control a dynamic system. On the other hand, Geddes's experiment suggested that goals have a significant effect on the quality of practice. If the finding of Geddes's experiment can extend to other control tasks, the same effect will probably also happen in the medical decision task of Harvey's studies. That is, the presence of different goals will have a significant effect on learning to control the task with consequent effects on the ability to make predictions (i.e. to forecast future situations). However, since there are critical differences between the two tasks, it is still too early to make precise predictions. In Harvey's studies, in addition to knowing that people have the ability to control the medical decision task, different effects on the ability to control and the bias in controlling the task were also investigated. It was found that there are significant differences in the degree of overconfidence between the executors and observers, but no in their control

performance. In addition, Geddes' observation experiment also suggested that there was no significant difference between the performance of observer group and executor group as long as they shared the same goal.

Chapter Two

Experiment One

EXPERIMENT ONE

The present study does not examine forecasting in a specific professional field. Instead, we are looking at how forecasting in general is affected by the conditions of learning. We will focus on the effect of goal on the medical control task and make a comparison with Geddes' results of similar research on control tasks. In this experiment, goal is the major independent variable to be manipulated. It is believed that by inducing different goals for the participants of the experiment, different effects on learning to forecast the behaviour of the system will result.

The study will provide a deeper understanding of human forecasting. Since the medical control task consists of different types of pattern, we can examine whether or not all the patterns are learned in the same way. It is worth seeing whether there are any interaction effects between the learning goals and the complexity of the control task on the use of rule-learning or instance learning. In the experiment, different goals will be manipulated for the participants and the process of learning will be examined. One goal will be problem solving: participants have the goal of providing treatment for the patients so as to bring their condition under control, while the other group has the goal of rule searching: they have to search for the pattern of the illness to provide information for research. The finding of the present studies will have important implications for professional decision making. There are many kinds of social research in society for social, market and medical purposes. Among them, different methods of investigation have been used. In some social research, the executor is asked directly to

describe the situation problems. In other research, a team of researchers are set up to investigate the problems. These methods induce different goals for the participants, if the effects of different goals are understood, there will be better use of the research on forecasting.

In the following experiment, we are going to investigate the behaviour of forecasting and test whether rule-learning will improve forecasting. The presence of a rule-learning effect is examined by setting task goals for the participants in the experiment. Instead of producing some well-known item for forecasting, we required the participant to face something new to them so that the outcome of the result will not be affected by their previous experience. Our intention is to observe the effect of the manipulated goals on the performance of the two groups: psychiatrists and researchers.

In modelling the treatment in the experiment, the logistic map formula is used. It is a remarkably suitable formula in the present experiment. First, it produces a variety of patterns for investigation. The logistic map has several important characteristics: its ubiquity, simplicity and stability and the different behaviours that it produces for different values of parameter A. Second, the logistic map formula is one of the chaotic formulae that are widely used in forecasting. Iterations of the particular chaotic formula associate with certain patterns of phenomena found by using time-series analysis. In recent years, they have been used to characterise certain unpredictable economic behaviours and revived many classical theories which typically over-simplify the real situation in the economy. The third reason is that mood disorder exhibits unpredictable behaviour which seems to be similar to a chaotic pattern.

Although it has not been proved that mood disorder can be described by a chaotic formula, more and more evidence has been found that non-linear methods can provide a better understanding of mood disorder (Paulus, Geyer & Braff 1996).

In this experiment, the goal effect on learning was examined and whether the manipulation of goals imposes an effect on subsequent forecasting. If there are some effects, what effects are they and to what extent do they affect cognitive processes and influence people's judgements and decisions? Since there exist substantial differences between the Clegg computer task of Berry and Broadbent's experiment and the medical decision task in the present experiment, it may not be possible to generalise Geddes' findings on the effect of goal to the present experimental task. Whatever, it is worth exploring the possible effects produced by goals. On the one hand, since human cognition is a complex process, we need to test any proposed generalisations of new experimental findings. On the other hand, inconsistent findings between similar but different experiments can usually provide deeper knowledge of the phenomenon and better understanding of cognitive processes. By this way, we will know either the actual general principles of cognition or what more specific variables have to be considered in formulating psychological theories.

If, in the present experiment, manipulation of goals produces a similar effect as in the Clegg task, participants in the psychiatrist group would tend to learn instances while participants in the research group would tend to learn rules. Thus, it is expected that participants in the research group will report the pattern of the illness much better than participants in the psychiatrist group. In addition, the performance of new

prediction cases by the research group should be better than that of the psychiatrist group. However, in case the effect of goals cannot extend to the present experimental task, there are several possible reasons. First, if the experiment provides reverse results, it might be due to the differences between executor-observer. In the experiment, only executors make decisions and they can get the outcomes of their own decisions while observers receive none. Furthermore, if there is no significant difference between the two groups of participants, it may be a matter of task difficulties. That is there may be some interactions between choice of learning patterns and the difficulty of the task.

The experimental task in the present experiment was quite similar to the medical decision task in Harvey's (1990) experiment. A logistic formula was used to characterise the pattern of patients' mood disorders. One group of participants acted as psychiatrists and they had to meet patients and administer drugs for them. Another group of participants acted as observers and played the roles of researchers who had to analyse the pattern of the illness and report it to the health authorities. The purpose of this setting of participants was to manipulate two different types of goals for both groups. The first group (called psychiatrist group or executor group) was designed to have the "specific" goal of problem solving while the second group (called research group or observer group) was designed to have the "non-specific" goal of rule searching.

Method

Participants

There were sixty voluntary participants, most of them being undergraduate students in psychology department of University of Durham. The remainder were postgraduates or staff in that department.

The Task

The control task in the experiment is composed of two elements. One is the behaviour of the system and the other is the mock situation. The system behaviour is characterised by a logistic map formula:

$$Y_{t+1} = A\, Y_t\, (\, 1 - Y_t\,)$$

The use of a logistic map formula to produce various patterns of behaviour provides a rich method for investigating the psychology of forecasting. It poses enough complexity to generate unpredictable situations for participants; on the other hand, it is not too complicated to secure meaningful observation. The system generated by a logistic map formula is called a non-linear dynamic system or chaotic system. Theories of these kinds of chaotic systems have long been discovered in the natural sciences. Chaos theories are described as the third scientific revolution of this century and are widely used in some areas of scientific forecasting (Lindsay 1997).

The above formula was iterated by a computer. For each iteration, the present output of the formula was set as the next input of the formula. Thus, after a number of iterations, a sequence of output values was produced. If all the output values are almost the same, the system is said to be stable. If the output values are not the same, the system would be under several possible states. First, the system may be under an alternative state, in which it alters between two values. The system produces a first output and a second output. The third output will have a similar value to the first and the fourth to the second, etc.. Second, the system may be under the several-value discretion state, in which it changes among several fixed values. That is the system undergoes periodic behaviours or a stable oscillation. The last one is the chaotic state; the system in this state will have an unpredictable changing pattern. It produces a series of outputs without following any apparent rules. Whereas the system is produced by the logistic map formula, the states of the system are determined by the value of parameter "A" in that formula. For A greater than 1.0 but less than or equal to 3.0, Y has a single stable value which increases with the value of A. When the value of A is above 3.0 but less than or equal to 3.4, Y alternates between two fixed values. As A increases above 3.4, Y moves between four values; as A increases further, Y moves between eight values. This doubling continues until A increases up to 3.57. For A greater than this value, the system produces unpredictable chaotic behaviour for most values of A.

Apart from the use of the logistic formula to characterise the output of the system of the experiment, a situation is introduced. The situation is a mock medical-decision situation in which a participant faces and treats several patients with mood

disorder. Participants are told explicitly the mock situation they are facing. Information about the nature of the mood disorder, the use of drugs, the characteristics of the treatment and the concept of a happiness index of the task are explained to them.

The purpose of using a chaotic formula to produce a mood disorder pattern is not to imitate real clinical cases but to enable observation of participants' response when facing chaotic behaviours. This produces a means for observing the human performances of prediction and judgement under the chaotic situation. Also, it provides participants with a real situation and thus increases the validity of the experiment. The original logistic map formula can only produce output values between zero to one. To suit the drug treatment situation, the formula has been modified so that the output value is multiplied by 50. The formula has been modified as follows:

$$Y_{t+1} = A\, Y_t \,(1 - Y_t /50)$$

In the mood disorder situation, changes in the happiness index will indicate patients' type of mood disorder (depressive, manic or manic-depressive). This is done by varying the starting parameter A of the modified logistic map formula in a computer program. Participants will initially be presented with the system with several possible states in which the behaviour of Y is constant (depressive or manic), alternating or chaotic (manic-depressive). In the experiment, the target stable value of output Y is 29-31. This will happen when A is between 2.4 and 2.7. For A below 2.4, the output Y will be a stable value lower than the target. It would be considered as a depressive case. For A between 2.7 and 3.0, the output Y will be a stable value higher than the

target. It would be considered as a manic case. Since output Y will oscillate for A greater than 3.0, it would be considered a manic-depressive case. Therefore, what the participants have to do is to alter the value A and make Y remain constant within the target values.

The task is completed by means of prescribing appropriate dosages i.e. a suitable amount of lithium or anti-depressant. For each unit of lithium dosage, the value of A decreases temporarily by 0.05. On the other hand, each unit of anti-depressant increases the value of A temporarily by 0.05. Since different patients have different starting A values that represent their types of mood illnesses, participants have to decide which drugs and how much they have to prescribe. Eight different starting A values are used (2.0, 2.2, 2.9, 3.1, 3.3, 3.5, 3.7 and 3.9) for eight different patient cases, which will be presented to each participant in a different random order. The task of the participant will be to alter the values of A in the eight different cases so that they fall between 2.4 and 2.7. On the other hand, the happiness index is represented by Y in the formula. In each case, output Y is obtained by successive iterations except the first one. The first value of Y is randomly selected from numbers between one to forty-nine. To remove any instability or transience of the formula, 100 self-iterations are performed before the start of a participant's trials.

Apparatus

The experiment was carried out on a personal computer. Each participant sat in a quiet room in front of the computer screen. Stimuli appeared on the screen and

they made responses by the computer keyboards.

Design

Participants were randomly allocated to one of two groups, a treatment goal group (psychiatrist group) and research goal group (research group). Both groups worked under eight different conditions which corresponded to the eight different starting values of A. There were eight trials in each of the eight conditions. The psychiatrist group were required to complete the eight conditions while the participants in the research group were required to observe the treatment given by the psychiatrist in each condition. The sequence of the conditions were given in different random orders. Each participant in the research group was paired with one participant in the psychiatrist group so that participants of the research group could observe, through computer data files, the whole series of treatments by the corresponding participants in the psychiatrist group.

Procedure

The experiment was divided into four parts. The first part consisted of treatment learning trials for the psychiatrist group and observation learning trials for the research group. The second part consisted of questionnaires in which participants were asked to describe the rules behind the behaviour of the mood disorders and rate the confidence of their descriptions. The third part consisted of prediction trials. Participants were asked to predict the next happiness index of the patients. The last

part consisted of judgement trials in which participants were asked to judge what dosages were suitable for the patients. Both groups were given written instructions of the experiment, which described their responsibilities in various parts of the experiment, gave simple descriptions of the illnesses and their corresponding treatment, the nature of the drugs etc.. These instructions were left beside the computer so that at any time, the participants could refer to the instructions if necessary. The instructions were as follows:

"Illness and Drug: Previous findings show that some people are consistently more depressed than normal people in the absence of any treatment. To control their mood states, these patients must be treated with an antidepressant drug. The more depressed they are, the more antidepressant must be prescribed to bring their moods into the normal range. Other patients with mood disorders are either manic or manic-depressive. Manic are consistently more irresponsible and boisterous than normal people. Manic-depressives' moods change between excessively boisterous and excessively depressed. In the absence of treatment, some manic-depressives' moods oscillate between the two extremes, whereas others have quite unpredictable changes in mood. Both manic and manic-depressive patients must be treated with the drug lithium.

"Characteristics of the treatment: Patients with mood disorder need continuous treatment. If at any time, the treatment stops, the patient will revert to the same pattern of moods as before any drugs have been prescribed. In addition, the dosage may need to be maintained for longer than a single month for it to bring the patient right into the

normal range of moods. Also, a consistently depressed patient maintained on an overdose of antidepressant will become manic or manic-depressive. Similarly, a consistently manic or manic-depressive patient maintained on an overdose of lithium will become depressed.

"Happiness index: The patients have been asked to keep a diary of experiences that make them feel noticeably happy. The total number of those experiences in a given month is known as the happiness index for that month. It is known that normal people usually have a monthly happiness index of between 29 and 31. Depressed patients have happiness indexes lower than normal while Manic patients have happiness indexes higher than normal. Manic-depressives' moods change either between high and low indexes or in unpredictable ways."

They were also given a summary table for the characteristics of the treatment as follows:

Patient types	**Depressed**	**Manic**	**Manic depressive**
Happiness indexes	Less than 29	Higher than 31	Unstable
Drugs used	Antidepressant	Lithium	Lithium

In addition to the above instructions, there were specific instructions for each group. The psychiatrist group was given the following instruction as the task for the experiment:

"Thank you very much for your participation. In this experiment, you are going to interact with the computer. The computer will present you with information about psychiatric patients with mood disorders. Your task is to prescribe drugs for those patients and try to bring each patient's mood back into the normal range. After you have done this prescribing task, you will see some treatments of patients and you will be asked to predict the changes of patients' conditions. Finally, you will be given some information about some patients and asked to give drug dosages that will bring their mood into the normal range."

For the research group the description of the experimental task is as follows:

"Thank you very much for your participation. In this experiment, you are going to interact with the computer. The computer will present you with information about a psychiatrist treating several patients with mood disorders. You are requested to play the role of medical researcher. Your task is to understand the pattern and rule behind this type of mood disorder illness so that you can report it to the research group of a health authority. After you have finished seeing the treatment record, you will see some treatments of patients and you will be asked to predict the changes of patients' conditions. Finally, you will be given some information about some patients and asked to give drug dosages that will bring their mood into the normal range."

Treatment trials:

In each 'A' value condition, a psychiatrist group participant saw two months' values of a patient's happiness index, after that they chose a drug and decided the suitable amount for the patient. They then saw the happiness indexes for the next eight months and administered a drug dose after each one, except the last. This completed one of the conditions. There were eight conditions in the experiments. Each condition started with a different starting value of A (2.0, 2.2, 2.9, 3.1, 3.3, 3.5, 3.7 and 3.9) in the logistic map formula. Each psychiatrist group participant carried out all eight conditions and the conditions were presented in different random order for each participant.

In these trials, each participant first saw the instruction on the computer screen: "Thank you very much for your participation. In this experiment, you are going to meet several patients with mood disorders. You will now be shown the first patient's two month mood index and you have to prescribe a drug for this patient." Then, they were asked to press the space bar on the keyboard to continue.

After that, it showed on the screen:

The first two months record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
1	None	32
2	None	24

They would then be asked to choose a drug and then prescribe the amount of drug for the patient. On the screen, it showed the following question just below the above

happiness index table:

Which drug would you like to prescribe?

Please type '1' or '2'

1. Lithium

2. Antidepressant

Participants then made their choice of drug by pressing the key of either '1' or '2' on the keyboard. The selected drug would then be printed on the screen and they would be asked to decide the amount of drug. For example, if a participant had selected lithium for this trial, the following question would be added on the screen:

What amount of Lithium (in mg) are you going to use?

Please type in a whole number between 0 and 30

Amount (mg):

The participant was required to press in a whole number between zero and thirty in order to complete the drug dose. The drug dose combined with the starting value of A produced the A value of next iteration which was put into the logistic map formula calculation. The result of this iteration produced the happiness index for the next month. This information would be added to the happiness index already on the screen. On the screen, it showed:

The last three months' record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
1	None	32

2	None	24
3	30 mg Lithium	28

Information from previous months scrolled up the screen so that only the last three months' information was shown. For example, if the fourth month information had been presented, the first month's information would not be shown. That means only the second, third and fourth month's information would appear on the screen:

The last three months' record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
2	None	24
3	30 mg Lithium	28
4	20 mg Lithium	32

Observation trials:

Participants in this group played the roles of researchers. On the screen, they were told that "Thank you very much for your participation. In this experiment, you are going to see a psychiatrist treating several patients with mood disorders. You will now be shown the first patient's two month mood index and then the successive treatment of that patient by the psychiatrist." Information in these trials was extracted from the previous data file. Since participants in of the psychiatrist group and the research group were paired up, participant thirty-one (first "researcher") would see the trials of participant one (first "psychiatrist"). Participant thirty-two (second "researcher")

would see the trials of participant two (second "psychiatrist"), etc. The information was displayed on the screen three months at a time in the same way as it was for the psychiatrist group.

The Questionnaire:

In this session, both groups were asked to describe the pattern of the mood illness and rating the confidence they had for their descriptions. They were asked the description question of "Please feel free to try to describe the underlying rule(s) behind the characteristics of the mood illness." and the confidence rating question "To what extent do you think that your description is correct? Please circle the choice." with the choice of "Very confident", "Confident", "Some confident", "Not confident" and "Really not sure". This questionnaire was presented on a separated sheet which was put with cover paper on the desk beside the computer. Participants were requested to answer the questions on that sheet.

Prediction trials:

All participants were given two successive happiness indexes and drug doses prescribed for a patient. They were asked to predict what the next happiness index would be if a specified amount of the relevant drug were prescribed to the patient. In the program, they were told "You are going to see some treatment cases. You will be

given the information the last two months from the patient. Then you will be shown the new treatment for the patient for this month. You have to predict what the happiness index will be at the end the that month's treatment" On the screen, participants were given information in the table formats. For example, participants might be asked:

The last two months' record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
1	None	24
2	6 mg Lithium	28

A drug dose of 18 mg Lithium is then given to the patient. What do you think the patient's happiness index will be at the end of the month? Please type in a number.

There were ten trials for each participant. Among these ten trials, five of them were old. That is, trials the participants had already seen in the treatment trials or observation trials. The data were extracted randomly from the treatment trial of the corresponding participant in the psychiatrist group. The other five prediction trials were new. Drug doses are integers that are randomly selected between zero and S ($S=20 \times (A_i - 2.5)$). That is the drug dose varied between zero and correct amount. The new predictions and old predictions were presented in different random orders.

Judgement Trial:

In this part, participants were asked to make judgements for the selection and dose of drug. They were given two successive months' happiness indexes of patients without

any treatment. Then they were asked to administer suitable drug doses. For example, on the screen, it shows:

The previous two month record of happiness index of a patient without any treatment is as follows:

Month	Drug Dose	Happiness Index
1	None	32
2	None	24

What dose of which drug do you think are suitable for that patient? (Lithium or Antidepressant, 0-30mg)

There were ten trials for each participant. The values of A_i and H were obtained randomly. A_i was selected randomly from the eight starting values. H was obtained by a randomly selected number between one and forty-nine following 100 self-iterations.

Results

Error was used as an index to measure the accuracy of participants' judgements. The smaller the errors were, the better the judgements were. In this experiment, absolute error was employed as a means to measure error. Both root mean square error and absolute error produced similar findings but the absolute error method was more useful and convenient than the method of root mean square error. Also, the absolute error would not be confounded by the summation of over-estimation and under-estimation as would the signed error. For simplicity, mainly absolute error was used in the present report. Nevertheless, when necessary, signed error is also mentioned to show error direction.

Prediction trials

Table 2.1 shows the mean absolute error in both old and new situations for each participant group and each condition. Errors were based on the happiness index in the prediction trials. The absolute error was the difference between participants' predictions and the correct values - the values generated by logistic formula iterations. Conditions were defined by the eight starting As and the mean absolute errors were the averages of the absolute errors of the participants in each group and each condition.

For example, in a condition where the starting value of A was 3.1; a participant had to give about 12mg Lithium in order to keep the A value of the patient dropping to

the optimal value of about 2.5.

The following table shows a record of the treatment of the participant:

Month	Value of A	Drug used	Happiness Index
1	3.1	-	16
2	3.1	-	34
3	2.9	4 Lithium	32
4	2.7	8 Lithium	31
5	3.1	0 Lithium	36
6	2.5	12 Lithium	25
7	2.5	12 Lithium	31
8	2.6	10 Lithium	30
9	2.6	10 Lithium	31
10	2.5	12 Lithium	29

The value of each Happiness Index was iterated by the formula by using each of the value A in the above table.

In the old cases of the prediction trials, the computer would randomly select his/her data in the treatment trials. If the computer selected the data of Month 6, 7 and 8, the participant would see the information as below:

The last two months' record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
1	12 mg Lithium	25
2	12 mg Lithium	31

A drug dose of 10 mg Lithium is then given to the patient. What do you think the patient's happiness index will be at the end of the month? Please type in a number.

The prediction error is the difference between participant's predicted value and the actual value. In the above case the actual value is 30. The Z value is the difference between the starting value of A. In this condition the starting A is 3.1 and the optimal value of A is 2.5 (in any condition). Hence, the value of Z is 0.6.

The analysis was divided into two stages. In the first stage, mean absolute errors of trials in each situation for each participant were analysed. First, a simple ANOVA with repeated measures was used to investigate the group effect, the situation effect and their interaction. It was found that there was no significant difference between the groups ($F(1, 58)<1$). However, the difference between old and new situations was marginally significant ($F(1, 58)=3.88$, $p=0.054$), the errors were greater in new situations than in old situations. The interaction between group and situation was not significant ($F(1, 58)=2.62$, $p=0.111$). However, since a specific prediction was made about performance on old and new situations by the two groups, the situation effect was further analysed in each group separately by paired sample t-tests. It was found that there was a significant situation effect in the psychiatrist group ($t(29)=2.66$, $p=0.013$) but not in the research group ($t=0.24$, $df=29$). The psychiatrist group performed better in the old situations than in the new situations.

Table 2.1: The mean absolute error in both old and new situations for each participant group and condition in the prediction trials.

	Conditions								Overall Means
	1	2	3	4	5	6	7	8	
Old situation									
Psychiatrist group	1.60	3.00	4.11	5.38	4.45	4.06	5.35	11.88	4.85
Research group	1.60	0.67	6.18	3.30	5.82	7.44	6.90	10.60	5.43
New situation									
Psychiatrist group	3.29	2.88	3.50	5.00	7.26	8.55	9.89	12.79	6.73
Research group	3.00	1.65	1.22	4.89	5.76	5.94	11.90	10.59	5.60
Means	2.41	1.99	3.49	4.59	5.77	6.66	8.52	11.50	5.65

In the second stage, the absolute errors were decomposed into each of the eight conditions. Since, the initial purpose of these trials was only to have a general comparison across old and new situations, only ten trials, with five old and five new, had been given to each participant. Hence each participant had three missing data points in both the old and new situations. Furthermore, the prediction trials were selected at random from the learning trials, so different cells were empty for different participants. Thus the data in Table 2.1 could not be analysed in a single ANOVA. Consequently, the comparisons between conditions were analysed using independent t-tests. The specific participants entering into each comparison varied according to whether or not the participants had data in each pair of conditions being compared. For example, to compare condition one across groups, only those subjects who had data in condition one could be used. Similarly, differences between old and new situations in

each condition were analysed using paired t-tests and once again, the participants entering into each comparison varied according to the available data points in old and new situations in each condition. As expected, there were significant condition effects in the experiment.

Participants performed significantly differently in the eight experimental conditions. Roughly speaking, there were increasing difficulties from condition one to condition eight with the exception that condition one was more difficult than condition two but less difficult than condition three, four, five, six, seven and eight. However, contrary to expectation, when the old and new situations were analysed separately, there was a significant group effect in the old situation of condition two ($t(20.73)=3.06$, $p=0.006$) and the new situation of condition three ($t(17.14)=2.65$, $p=0.017$); the research group had a smaller error than the psychiatrist group in these two cases. This only happened in two out of sixteen (two situations X eight conditions) cases. It seemed not to be a Type I error since the probabilities were very low. A significant difference between old and new situations was only found in condition seven of the research group with ($t(14)=2.35$, $p=0.034$). It was believed that it was a Type I error, since it was found in one out of sixteen (two groups by eight conditions) cases. Thus, it was concluded that situation effect in each group by each condition was not significant. By observing the signed error, there was no preferred direction to the error - overshooting or undershooting of the estimated happiness index. The raw data of signed errors is in Appendix 1.1 and Appendix 1.2.

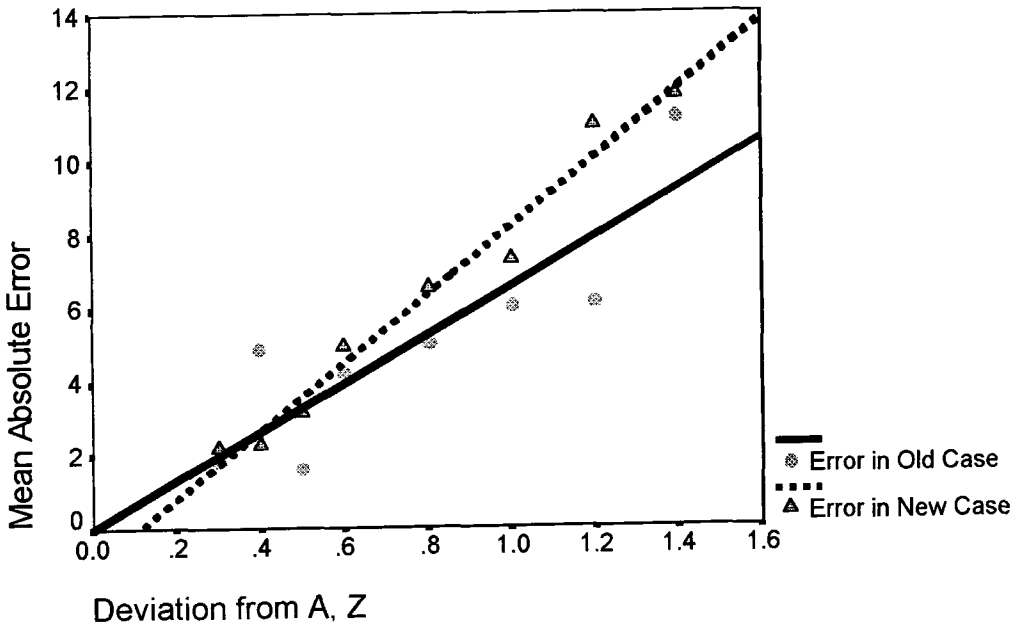
Chaotic behaviour in the prediction trials

An important discovery in the experiment was the unexpected high correlation between participants' errors on the prediction task and the parameter of the logistic map formula. In examining the correlation between the two variables, absolute mean error and the deviation from the optimum A value were compared. For convenience, deviations from the optimum was represented by Z here. In the experiment, the eight conditions had eight different values. The A value of normal people was 2.5 and the deviation from 2.5, Z, was used to compare with the absolute errors. As mentioned in the method section, the A values in the eight conditions were 2.0, 2.2, 2.9, 3.1, 3.3, 3.5, 3.7 and 3.9 respectively, thus, the Z values were 0.5, 0.3, 0.4, 0.6, 0.8, 1.0, 1.2 and 1.4 respectively. The data were analysed with Pearson's correlation. The correlation between the absolute error and Z was R=0.9878 (p=0.000) for the new situations of the prediction trial and R=0.8697 (p=0.000) for the old situations. Further analysis revealed that R=0.9896 for the psychiatrist group and that R=0.9465 (p=0.000) for the research group in new situations; R=0.7549 (p=0.000) for the psychiatrist group and that R=0.8500 for the research group in old situations.

Table 2.2: The values of A and Z and mean absolute error in both old and new situations for each condition in the prediction trials.

	Conditions							
	1	2	3	4	5	6	7	8
Values of A	2.0	2.2	2.9	3.1	3.3	3.5	3.7	3.9
Values of Z	0.5	0.3	0.4	0.6	0.8	1.0	1.2	1.4
Mean absolute error in the old situation	1.60	1.80	4.90	4.22	5.05	6.07	6.15	11.19
Mean absolute error in the new situation	3.16	2.18	2.29	4.95	6.56	7.32	10.95	11.75

Figure 2.1: Plot of mean absolute error against A deviation, Z in the prediction trials



Judgement trials

Table 2.3 shows the mean absolute error for each participant group and each condition in the judgement trials. In these trials, the absolute errors were measured by drug dosage in millilitres not the happiness index as in the prediction trials. The data in this table were analysed using a two (group) by eight (condition) ANOVA with repeated measures on the last factor. There was no effect of group ($F<1$) but there was a significant effect of condition similar to that in the prediction trials ($F(7, 406)=78.27$, $p<0.001$). The interaction between group and condition failed to reach significance at the 0.05 level ($F(7, 406)=1.79$, $p=0.087$). One way ANOVAs were used to look into the group difference in each condition, there was a significant group difference in condition two ($F(1, 59)$, $p=0.0312$) the research group had significantly less error than the psychiatrist group. It was believed to be a Type I error based on two reasons. First, it was found in one out of the eight conditions; second, the opposite trend were found in its two adjacent conditions. Table 2.4 shows the mean signed error for each participant group in each condition in the judgement trials. Negative numbers in this table meant under-dosage of drugs. It was found that participants apparently had a general tendency for under-dosage.

Table 2.3: The mean absolute error for each participant group and condition in the judgement trials.

	Conditions								Overall Means
	1	2	3	4	5	6	7	8	
Psychiatrist group	4.37	4.20	5.43	6.87	8.03	9.67	17.97	22.27	9.85
Research group	5.80	2.67	5.90	7.93	10.60	13.03	15.50	24.93	10.80
Means	5.09	3.44	5.67	7.40	9.32	11.35	16.74	23.60	10.32

Table 2.4: The mean sign error for each participant group and condition in the judgement trials.

	Conditions								Overall Means
	1	2	3	4	5	6	7	8	
Psychiatrist group	-3.57	-3.47	-5.10	-6.33	-6.97	-9.67	-17.97	-22.13	-9.40
Research group	-5.00	-2.40	-4.63	-7.53	-9.27	-12.70	-15.37	-24.93	-10.23
Means	-4.29	-2.94	-4.87	-6.93	-8.12	-11.19	-16.67	-23.53	-9.82

Treatment trials (learning trials of the psychiatrist group)

Table 2.5 shows the mean absolute error in each condition of these learning trials. In these trials, learning was investigated in addition to the condition effect. The eight trials in each participant's eight conditions were divided into two. The first four trials were compared against the last four trials for each condition. The data in this

table were analysed using a two (order) by eight (condition) repeated measures ANOVA. As expected, errors in the last four trials were significantly lower than in the first four ($F(1, 29)=6.58$, $p=0.016$). A significant condition effect was also found in general similar to that in the prediction trials ($F(7, 203)=45.36$, $p<0.001$). There was no significant interaction between condition and order ($F(7, 203)=0.51$, $p=0.828$). Also, participants had a general tendency of under-dosage as in the judgement trials.

Table 2.5: The mean absolute error for each condition in the learning trials.

	Conditions								Overall Means
	1	2	3	4	5	6	7	8	
First four trials	4.93	4.18	6.38	7.30	9.46	12.39	14.04	21.28	10.00
Last four trials	3.28	3.83	5.92	6.67	8.73	12.27	13.27	19.62	9.20
Means	4.11	4.01	6.15	6.99	9.10	12.33	13.66	20.45	9.60

Questionnaire

In the questionnaire section, it was found that participants could not give any meaningful description of the underlying rule(s) behind the characteristics of the mood disorders. Most of the participants just repeated the basic ideas of the illness given in the experimental instructions. For example, one participant gave a description that "the more the patients happiness index varies from normal the more treatment they require". Another participant gave the description that "Depressive people given an

anti-depressant become overly happy and unstable. Lithium given in manic-depressive can control. Either the people have a fairly stable range of happiness or they can range between the extremes - manic depressive." Some of the participants just described their observation without any analytical description of the rules behind. For example, one response was "If the patient has an alternating happiness index then a problem appears to exist". Another problem is that participants seemed to have different ideas or perceptions and corresponding answers when asked to describe the rule behind. Some seemed to think that they were asked to describe their observation and some describe the difficulties they encountered. Thus, no observable differences between the two groups could be observed from these responses. For the confidence rating of their description, an independent t-test was used to analyse the results. No significant between group effects ($t=0.15$, $df=58$) could be found. There was no significant correlation between confidence ratings with performance in either the judgement trials or the prediction trials.

Group correlation

For both prediction trials and judgement trials, the performance of the participants in the psychiatrist group were compared with the matched observers in the research group. The data were analysed with Pearson's correlation; some degree of correlation was found. In the prediction trial, $R=0.4157$, $p=0.022$ two tailed. In the judgement trial, $R= 0.3803$, $p=0.038$, two tailed.

Discussion

In general there was no significant group effect in the prediction trials, nor a significant interaction between group and situation. A situation effect in the prediction trials was marginally significant, while a significant difference between old situation and new situation was found in the psychiatrist group but not in the research group. By comparing the effect of goals in this experiment with a previous experiment (Geddes and Stevenson 1997) on a dynamic control task (Berry & Broadbent 1984), there are similarities as well as differences between them. In Geddes' experiment, it was found that in the prediction trials, there was significant group effect as well as situation effect but no significant interaction between group and situation, while a significant difference between old situation and new situation was found in the specific goal group but not in the non-specific goal group. The non-specific goal group in Geddes' experiment had a non-specific goal of rule searching as the research group in the present experiment while the specific goal group in Geddes' experiment had a specific goal of problem solving as the psychiatrist group in the present experiment. Whenever comparisons are made in the present discussion, both the non-specific goal group in Geddes' experiment and the research group in the present experiment are referred to as the rule searching group and both the specific goal group in Geddes' experiment and the psychiatrist group in the present experiment are referred to as the problem solving group.

The result of a non-significant group effect is not consistent with the findings of Geddes' experiment. According to Geddes, the poorer performance of his problem

solving group relative to his rule searching group was because the problem solving group had only learned instances whereas the pattern searching group had learned rules. There may be several reasons to explain the conflicting results. First, in the control task of Geddes' experiment, the response required a qualitative decision but not a quantitative judgement, as in the medical decision task of the present experiment. The qualitative task is a relatively easy one compared to the quantitative task used in the present study. It is possible, therefore that the Geddes' problem solving group had little motivation to search for the rule because the task could be learnt without having to do so. By contrast, the problem solving group in the present study may have been sufficiently motivated by the difficulty of the task to search for rules. Alternatively, the present task may have been too difficult to learn implicitly resulting once again in rule learning on the part of the problem solving group. According to both of these interpretations, the lack of difference between the two groups arises because the problem solving group as well as the rule searching group learned rules. However, since the problem solving group were better at predicting from old compared to new situations, indicating that they have learned instances, this possibility seems unlikely. The alternative, and probably more likely, possibility is that the complexity of the task prevented the rule searching group from fully grasping the underlying rule. the results of the prediction tests suggest that this group had some grasp of the rule, since they could predict successfully from both old and new situations. However, it is likely that they did not fully understand the rule, given the complexity of the task and the small number of learning trials.

About the significant difference between old situation and new situation in the

psychiatrist group but not in the research group, this finding is consistent with the previous experiment by Geddes. The difference may be explained by the fact that psychiatrist group tended to have more instance learning and this enhanced its performance in the old situations but less rule learning and this impaired its performance in the new situations. However, there is another possibility. In the experiment, only the psychiatrist group could control the system. Implicit memory or instance learning might be caused by the effect of control. With this significant situation effect in the psychiatrist group, it could be inferred that implicit memory is also possible in the quantitative case.

In this experiment, since participants were not required to complete every condition in each situation, the significance of the interaction between group and situation could only be tested in all eight conditions combined and with each participant having a different combination of five out of the eight conditions. Because of this difficulty with the data, in further analysis of the prediction trial error, the data were pooled across conditions and independent t-tests examined differences between the subject groups while paired t-tests examined differences between old and new situations. This is not a typical method to analyse experimental data. Although the findings may still be reasonably reliable, it is better to have a full ANOVA repeated measures analysis with all participants taking part in all the conditions. This will be done in the next experiment.

About the unexpected significance of group in the old situation of condition two and the new situation of condition three, if these findings are reliable what do they

mean? If the research group can perform better than the psychiatrist group in simple conditions, the same thing should have happened also in the old situation of condition three and the new situation of condition two. Perhaps, there were some subtle interactions between situation and condition. However, with the original experiment design, this cannot be analysed by typical statistical tests. On the other hand, if the these findings are not reliable, what is the source of error? Although the statistical method is not typical, it still is a fair comparison across the two experimental groups. In examining the experimental data, the absolute errors of the participants, it is suspected that the selection of cases for the participants might have produced some undesirable effects. Trials in this part of the experiment were selected at random by the computer. Prediction trials for "psychiatrists" were not the same as those for "researchers". Since the difficulties of the trials varied even within the same condition, it was not possible to be sure that the difficulties of the trials were reasonably similar for both groups. Therefore, the findings in this part of the experiment might not be reliable. In the next experiment, the method will be modified in order to prevent the same problem from happening again.

In the prediction trials, it is found that absolute errors made by participants had a very high correlation with the parameter A in the logistic map formula.

$$Y_{t+1} = A\, Y_t\, (1 - Y_t)$$

The situation in the control task is that participants are required to keep the output Y within a small stable range 29-31. It can be done if A is kept around 2.5. If the starting

value of A is equal to 2.9, the difference between starting value of A and the desirable value of A is 0.4. The difference between the starting values of A in any condition with the desirable value of A 2.5 is the deviation of A, Z. The greater Z is, the greater the prediction error is. In the trials, participants were requested to guess the next output of the formula, the happiness index. They had to make decision based on the drug dosage. The drug dosage will temporarily change the value of A. For each millilitre of drug the value of A will alter by 0.05 Combined with the starting value of A, this produces a temporary value of A for the logistic formula. Despite the effect of drug dosage, the correlation between prediction error and Z is still very high in the experiment. Why this correlation is so high? Perhaps it is reasonable to think that the greater the parameter A, the more unpredictable the system will be. However, it seems to be difficult to explain why the correlation is so high. A high correlation seems to imply that the system is linear to a high extent. The linearity of the relationship is thus the originality of these experimental findings. In the experiment, participants are not just required to guess the next output generated by the logistic map formula only. Instead, they have to make their decision in the presence of the variation of the parameter A. The variation of parameter A is selected randomly by the computer (at least for new situations). Having high correlation between starting values of A and prediction errors is not an obvious finding.

Several possibilities to account for the present result have been considered. Perhaps, it might be that this outcome is built in to the design rather than a property of the human decision system.

The first possibility is that participants' input in some way would manipulate the iteration of the logistic formula and produce a high correlation between their prediction errors and their input values. However, this is impossible in the present experimental design. Participant in these trials would not have the chance to affect the iteration of the formula by changing the value of A. In the prediction trials, participants were presented the treatment information of patients in the past two months which included the amount of drug dosage received by the patients and the corresponding happiness index. Based on this information, they were asked to predict patients' happiness index in the next month. Unlike in the treatment trials, participants in the prediction trials were not asked to give any treatment or to control the task any more. The purpose of these trials was to test participants' ability in predicting outcomes of the next happiness index. Their input of the prediction of happiness index of the patients thus would not affect the value of A and it is impossible for the participants to have any manipulation to the iteration of the logistic formula.

Another possibility is that observation of the cases in these trials in some way affected participants' decision and caused the particularly high correlation. In the prediction trials, sixteen cases were presented. Half of them were retrieved randomly from the treatment or observation history of the particular participant for the old prediction cases. The other half of the cases were newly manipulated by the computer program for the new prediction cases. These sixteen cases were then randomised by the computer before being shown to a participant. It is believed that participants could not have any systematic observation in these trials.

It is also possible that memory of participants may have had an effect on the high correlation. However, by observing the result of the new prediction cases, in which participants have to make their prediction without previous observation or manipulation of the particular cases, a similar pattern of correlation was found. Thus, the possibility that this high correlation was due to participants' memory was also ruled out.

Up to the present, no cogent explanation for this high correlation can be concluded. In the next experiment, all participants will take part in all the conditions in the prediction trials. Hence, it can be observed whether the same findings can be obtained. In comparing the correlation across situation and group, it was found that the correlation is higher in the new situations than the old. This may be due to implicit memory and instance learning. In the old situation, since participants may have remembered some of the cases, they may simply answer the prediction question using the memory of those old cases. This would have contaminated the correlation of Z and prediction errors.

The implication of these experimental findings will depend on the extendibility of the findings. If the findings can be extended to other chaos formulae, the application of the findings will be very useful. If the chaos formula of a system is derived, the prediction errors can probably be predicted by the parameter of that formula. In financial systems, for example, prediction errors highly concern the risk of investment. If the prediction errors can be estimated more accurately, the risk of investment can also be better estimated. Thus, a better financial decision can be worked out. Similarly,

the findings may also apply in other social sciences to which chaos theory has been applied.

In the judgement trials, the two groups performed similarly, even though the research group participants did not have the chance to control the system. This result is consistent with findings in other control task experiments with executor and observer settings. In these trials, participants were given two successive months' happiness indexes of patients without any treatment and were asked to administer suitable drug doses for them. These judgement trials are similar to the original treatment learning trials. Hence they can be regarded as a test of learning and indicate that observers and actors learn equally well. However, the matched pairs from each group did not have the same trials. Hence, the judgement trials were not as well controlled as they might have been. In the next experiment, therefore, the trials will be more carefully controlled, as in the prediction trials, in order to get a stronger comparison of differences between human prediction and judgement.

The purpose of analysing the treatment trials is to examine the presence of learning. Since performance in the last four trials was significantly better than the first four trials in each condition, it is believed that participants had learned through the treatment trials. For the research group participants, since they performed similarly with their counterparts in the psychiatrist group in both the prediction trials and judgement trials, it is also believed that they had learned through the observation trials. All participants in both treatment trials and judgement trials had a general tendency of under-dosage. Perhaps, when participants were not sure of the suitable amount of

drug, they just gave a small amount to prevent over-dosage. Written description seemed not to be a good means for observation in the experiment. Participants could not write down meaningful descriptions of the rule behind the illness or treatment for between group comparison. Hence, no meaningful findings could be obtained in this part of the experiment. In the next experiment, instead of asking participants to give written description, multiple choice questions will be used.

The correlation between performance of members within matched pairs on prediction and judgement trials was about 0.4. The reason for determining these correlations is to assess the influence of performance of the executor on the performance of the observer. Performance of participants in the psychiatrist group were compared with the matched observers in the research group. The correlation of 0.4 means about 16% of observers' performance was contributed by actors' performance. This seems to be a reasonable but not really significant proportion. However, since participants have different trials in both prediction and judgement, the difficulty of trials within matched pairs may be different. The correlation therefore may not be very reliable. In the next experiment, the experimental design will be modified so that participants with the same matched pairs will see the same questions.

Finally, one further problem with this first experiment is that the roles of the participants (executors or observers) is confounded with the learning goals (problem solving or rule searching). This problem will also be addressed in the second experiment.

Chapter Three

Experiment Two

EXPERIMENT TWO

A few modifications to the experimental design will be made in the present examination. The aims of these modifications are to enhance the validity and reliability of experiment one as well as to get better observation and deeper comparison of the prediction and judgmental behaviour in the control task. First, to improve the validity of the experiment, a three group design will be employed instead of a two group design. Second, to improve the reliability, a new method of selecting both prediction trials and judgement trials will be used to control a potentially influential random reliable. Also, in both prediction trials and judgement trials, eight conditions will be used instead of five. Furthermore, the format of the judgement trials will be modified in order to make more meaningful comparisons with the prediction trials. Finally, multiple choice questions will be used in the questionnaire section in order to get more meaningful findings.

In the last experiment, two groups of participants were used. One group (the psychiatrist group) acted as psychiatrists and implemented their treatment. The other group (the research group) acted as researchers and observed their treatment. That means that in the experiment, executors had a goal of problem solving and observers had a goal of rule-searching. Thus, it is possible that differences between groups in experiment one were produced by the participant's role (executor or observer) rather than the learning goal (problem solving or rule searching). To solve this problem, in the present experimental design, one more group will be added to the experiment. The

new group, called the trainee group, will have a goal of problem solving and a role of observer (See Table 3.1). Therefore, if there exist differences between the psychiatrist group and the research group, the results of the trainee group can be examined. If the result of the trainee group is significantly different from the research group but not significantly different from the psychiatrist group, the differences between the research group and the psychiatrist group will be considered to be produced by goals. On the other hand, if the result of the trainee group is significantly different from the psychiatrist group but not significantly different from the research group, the differences between the research group and the psychiatrist group will be considered to be produced by roles.

Table 3.1: The three-group design in the present experiment

	Role of observer	Role of executor
Problem-solving goal	Trainee group	Psychiatrist group
Rule-searching goal	Research group	

Thus only the psychiatrist group participants will actually do the treatment. The other two groups will observe the treatment by the "psychiatrist". Each psychiatrist group participant will be matched with one participant in the research group and another in the trainee group. In each set of matched triples, the research group participant and the trainee group participant will observe the treatment by the psychiatrist group participant. There are some reasons for using two observer groups and one executor group rather than two executor group and one observer group. If two-executor-group-and-one-observer-group design is used, there will be two sets of

source data of the treatment trials in each matched triple. However, if two-observer-group-and-one-executor-group design is used, there will only be one set of source data of the treatment trials in each matched triple. In the latter case, all matched members across the three groups will observe the same information. In this way, it is believed that the comparison will be more reliable.

As mentioned in the discussion section of the last experiment, since the trials were selected randomly by the computer, trials for participants in the psychiatrist group were not the same as those for their corresponding participants in the research group. As the difficulties of the trials varied even within the same condition, it was not possible to be sure that the difficulties of the trials were reasonably similar for both groups. Therefore, the findings in that part of the experiment might have been contaminated. This situation occurred in both prediction trials and judgement trials. To improve the control of this random variable, the trials for psychiatrist participants will be given to their corresponding matched participants in the other two groups. The computer programme will be designed in a way that trials for each condition will be randomly selected for each psychiatrist participant and this particular set of trials will be recorded in the data file and then given to the other two participants in the triple. Thus in each triple, all participants will see the same trials in both prediction trials and treatment trials.

In the old design of the judgement trials, participants were asked to make a prescription given two months' happiness index of a patient without any treatment. The purpose of these trials was to compare the judgement abilities of the two groups

and further to infer the quality of learning manipulated by the two different goals. In the previous experiment, no significant difference could be found between the two groups. In the present experiment, it is intended to further examine these judgements by comparing the results of the judgement trials with those of the prediction. It will be done by modifying the judgement trials and making them logically similar to the prediction trials. In the judgement trials, instead of asking participants to make a prescription for a patient without any treatment, they will be asked to make a prescription for a patient with a certain previous treatment. Each participant will be shown a patient's happiness index and drug treatment of the previous two months, and will then be asked to make their judgement on the next prescription in order to control the patient happiness index. In addition, two kinds of situations will be introduced in the judgement trials. Some of the trials will be old and the others will be new. Old trials are the trials participants have seen in the learning trials (that is, treatment trials for the psychiatrist group or observation trials for the other two groups). The trials will be extracted from the data file and appear in the judgement trials as the old trials. New trials will be generated randomly by the computer programme for the psychiatrist group and the other two groups will see the same trials as the matched member in the psychiatrist group.

In the previous experiment, each participant saw five old trials and five new trials, and each trial was randomly selected from the eight treatment conditions. In the present experiment, however, in order to make more precise comparisons across the eight conditions, they will be shown eight old trials and eight new trials for each condition. This will be done in both prediction trials and judgement trials.

In the previous experiment, since no meaningful description of the underlying rule(s) behind the characteristics of the mood disorders could be made by the majority of the participants, no meaningful findings nor any sensible comparison across the two groups could be observed. It was believed that it was too difficult for the participants to understand the exact requirement of the question and also to describe any of the complicated rules behind the mood disorders. In the next experiment, it is intended to modify this section in order to make the questionnaire easier for the participants and make the findings clearer for comparison. Instead of asking descriptive questions, they will be given several multiple choice questions. They will be required to choose the best description of the rules behind the mood disorders.

Method

Participants

There were seventy two voluntary participants, all of them being undergraduate and postgraduate students in University of Durham.

The Task

The same task as experiment one was given to the participants.

Apparatus

The experiment was carried out on a personal computer. Each participant sat in a quiet room in front of the computer screen. Stimuli appeared on the screen and they made responses by the computer keyboards.

Design

Participants were randomly allocated to one of three groups, two treatment goal groups (the psychiatrist group and the trainee group) and one research goal group (the research group). All three groups worked under eight different conditions which corresponded to the eight different starting values of A. There were eight trials in each of the eight conditions. The psychiatrist group were required to complete the eight

conditions while the participants in the research group and the trainee group were required to observe the treatment given by the psychiatrist in each condition. The sequences of the conditions were given in different random orders. Each participant in the research group and the trainee group was matched with one participant in the psychiatrist group so that participants of the research group and the trainee group could observe, through computer data files, the whole series of treatments by the corresponding participants in the psychiatrist group.

Procedure

The experiment was divided into four parts. The first part consisted of treatment learning trials for the psychiatrist group and observation learning trials for the research group and the trainee group. The second part consisted of questionnaires in which participants were asked to identify the rules behind the behaviour of the mood disorders by answering multiple choice questions. The third part consisted of prediction trials. Participants were asked to predict the next happiness index of the patients. The last part consisted of judgement trials in which participants were asked to judge what dosages were suitable for the patients. Both groups were given written instructions of the experiment, which described their responsibilities in various parts of the experiment, gave simple descriptions of the illnesses and their corresponding treatment, the nature of the drugs etc.. These instructions were left beside the computer so that at any time, the participants could refer to the instructions if necessary. The instructions were as follows:

"Illness and Drug: Previous findings show that some people are consistently more depressed than normal people in the absence of any treatment. To control their mood states, these patients must be treated with an antidepressant drug. The more depressed they are, the more antidepressant must be prescribed to bring their moods into the normal range. Other patients with mood disorders are either manic or manic-depressive. Manic are consistently more irresponsible and boisterous than normal people. Manic-depressives' moods change between excessively boisterous and excessively depressed. In the absence of treatment, some manic-depressives' moods oscillate between the two extremes, whereas others have quite unpredictable changes in mood. Both manic and manic-depressive patients must be treated with the drug lithium.

"Characteristics of the treatment: Patients with mood disorder need continuous treatment. If at any time, the treatment stops, the patient will revert to the same pattern of moods as before any drugs have been prescribed. In addition, the dosage may need to be maintained for longer than a single month for it to bring the patient right into the normal range of moods. Also, a consistently depressed patient maintained on an overdose of antidepressant will become manic or manic-depressive. Similarly, a consistently manic or manic-depressive patient maintained on an overdose of lithium will become depressed.

"Happiness index: The patients have been asked to keep a diary of experiences that make them feel noticeably happy. The total number of those experiences in a given month is known as the happiness index for that month. It is known that normal people

usually have a monthly happiness index of between 29 and 31. Depressed patients have happiness indexes lower than normal while Manic patients have happiness indexes higher than normal. Manic-depressives' moods change either between high and low indexes or in unpredictable ways."

They were also given a summary table for the characteristics of the treatment as follows:

Patient types	**Depressed**	**Manic**	**Manic depressive**
Happiness indexes	Less than 29	Higher than 31	Unstable
Drugs used	Antidepressant	Lithium	Lithium

In addition to the above instructions, there were specific instructions for each group. For the psychiatrist group, it was given the following instruction as the task for the experiment:

"Thank you very much for your participation. In this experiment, you are going to interact with the computer. The computer will present you with information about psychiatric patients with mood disorders. Your task is to prescribe drugs for those patients and try to bring each patient's mood back into the normal range. After you have done this prescribing task, you will see some treatments of patients and you will be asked to predict the changes of patients' conditions. Finally, you will be given some information about some patients and asked to give drug dosages that will bring their mood into the normal range."

For the research group the description of the experimental task is as follows:

"Thank you very much for your participation. In this experiment, you are going to interact with the computer. The computer will present you with information about a psychiatrist treating several patients with mood disorders. You are requested to play the role of medical researcher. Your task is to understand the pattern and rule behind this type of mood disorder illness so that you can report it to the research group of a health authority. To help you to do this, you will first observe the psychiatrist's drug treatment and its outcome for each patient. After you have finished seeing the treatment record, you will see some treatments of patients and you will be asked to predict the changes of patients' conditions. Finally, you will be given some information about some patients and asked to give drug dosages that will bring their mood into the normal range."

For the trainee group the description of the experimental task is as follows:

"Thank you very much for your participation. In this experiment, you are going to interact with the computer. The computer will present you with information about a psychiatrist treating several patients with mood disorders. You are requested to play the role of medical trainee. Your task is to learn how to treat the patients yourself, by prescribing drugs for the patients and trying to bring each patient's mood back into the normal range. To help you to do this, you will first observe the psychiatrist's drug treatment and its outcome for each patient. After you have finished seeing the treatment record, you will see some treatments of patients and you will be asked to

predict the changes of patients' conditions. Finally, you will be given some information about some patients and asked to give drug dosages that will bring their mood into the normal range."

Treatment trials: (The psychiatrist group)

In each 'A' value condition, a psychiatrist group participant saw two months' values of a patient's happiness index, after that they chose a drug and decided the suitable amount for the patient. They then saw the happiness indexes for the next eight months and administered a drug dose after each one, except the last. This completed one of the conditions. There were eight conditions in the experiments. Each condition started with a different starting value of A (2.0, 2.2, 2.9, 3.1, 3.3, 3.5, 3.7 and 3.9) in the logistic map formula. Each psychiatrist group participant carried out all eight conditions and the conditions were presented in different random order for each participant.

In these trials, each participant first saw the instruction on the computer screen: "Thank you very much for your participation. In this part of the experiment, you are going to meet several patients with mood disorders. You will now be shown the first patient's two month mood index and you have to prescribe a drug for this patient." Then, they were asked to press the space bar on the keyboard to continue.

After that, it showed on the screen:

The first two months record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
1	None	32
2	None	24

They would then be asked to choose a drug and then prescribe the amount of drug for the patient. On the screen, it showed the following question just below the above happiness index table:

> Which drug would you like to prescribe?
>
> Please type '1' or '2'
>
> 1. Lithium
>
> 2. Antidepressant

Participants then made their choice of drug by pressing the key of either '1' or '2' on the keyboard. The selected drug would then be printed on the screen and they would be asked to decide the amount of drug. For example, if a participant had selected lithium for this trial, the following question would be added on the screen:

> What amount of Lithium (in mg) are you going to use?
>
> Please type in a whole number between 0 and 30
>
> Amount (mg):

The participant was required to press in a whole number between zero and thirty in order to complete the drug dose. The drug dose combined with the starting value of A

produced the A value of next iteration which was put into the logistic map formula calculation. The result of this iteration produced the happiness index for the next month. This information would be added to the happiness index already on the screen. On the screen, it showed:

The last three months' record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
1	None	32
2	None	24
3	30 mg Lithium	28

Information from previous months scrolled up the screen so that only the last three months' information was shown. For example, if the fourth month information had been presented, the first month's information would not be shown. That means only the second, third and fourth month's information would appear on the screen:

The last three months' record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
2	None	24
3	30 mg Lithium	28
4	20 mg Lithium	32

Observation trials: (The research group and The trainee group)

Participants in these two groups played the roles of observers. On the screen, they were told that "Thank you very much for your participation. In this part of the experiment you are going to see a psychiatrist treating several patients with mood disorders. You will be now be shown the first patient's two month mood index then the successive treatment of that patient by the psychiatrist." Information in these trials was extracted from the previous data file. Since participants in the psychiatrist group, the research group and the trainee group were matched up, the first participant in the research group and the first participant in the trainee group would see the trials of the first participant in the psychiatrist group. The second participant in the research group and the second participant in the trainee group would see the trials of the second participant in the psychiatrist group, etc. The information was displayed on the screen three months at a time in the same way as it was for the psychiatrist group.

The Questionnaire:

In this session, all three groups were asked to answer the multiple choice question regarding the rules behind the behaviour of the mood disorders. They were told that there were several hypotheses claimed by some researchers. They were asked to decide whether they agreed with their hypotheses. On the screen, it showed "Some researchers have the following hypotheses about the pattern of the illness. Please state whether you agree with the following statements." Then they were given following

four hypotheses about the behaviour of the mood disorders in different random orders:

1. "If depressed patients are given the right amount of anti-depressant their mood must be able to shift back to the normal range."
2. "Manic-depressive patients seem to have manic pattern when prescribed insufficient amounts of lithium."
3. "The patients seem to be get a permanent negative effect when prescribed improper drugs."
4. "Some manic-depressive patients have an uncontrollable pattern."

After displaying each of the above statements, they were given the choice of "1. Yes", "2. No" and "3. Don't know". They presented their choices by press the keys of "1", "2" or "3" on the keyboard.

Prediction trials:

All participants were given two successive happiness indexes and drug doses prescribed for a patient. They were asked to predict what the next happiness index would be if a specified amount of the relevant drug were prescribed to the patient. In the program, they were told "In the next part of the experiment, you are going to see some treatment cases. You will be given information about the treatment of the patient for the last two months. Then you will be shown the new treatment for the patient for this month. You have to predict what the happiness index will be at the end of that

month's treatment." On the screen, participants were given information in the table formats. For example, participants might be asked:

The last two months' record of the patient's happiness index is as follows:

Month	Drug Dose	Happiness Index
1	None	24
2	6 mg Lithium	28

A drug dose of 18 mg Lithium is then given to the patient. What do you think the patient's happiness index will be at the end of the month? Please type in a number.

There were sixteen trials for each participant, eight of them were old. That is, trials the participants had already seen in the treatment trials or observation trials. The data were extracted randomly from the treatment trial of the corresponding participant in the psychiatrist group. The other eight prediction trials were new. Drug doses were integers that were randomly selected between zero and S ($S=20 \times (A_i - 2.5)$). That is the drug dose varied between zero and correct amount. The new predictions and old predictions were presented in different random orders.

Judgement trials:

In this part, participants were asked to make judgements for the selection and dose of drug. They were given two successive months' happiness indexes of patients and the treatments they received in these two months. Then they were asked to administer suitable drug doses. For example, on the screen, it shows:

The last two month's record of the patient's happiness index is as follows:

Month	Drug type	Drug dose	Happiness index
1	Lithium	8	12
2	Lithium	5	15
3	?	?	

What type of drug do you think should be given to this patient in month 3?

There were sixteen trials for each participant. The values of A_i and H were obtained randomly. A_i was selected randomly from the eight starting values. H was obtained by a randomly selected number between one and forty-nine following 100 self-iterations.

Results

Prediction trials

Table 3.2 shows the mean absolute error in both old and new situations for each participant group and each condition. Errors were based on the happiness index in the prediction trials. The absolute error was the difference between participants' predictions and the correct values - the values generated by logistic formula iterations. The data in this table were analysed by using a three (group) by two (situation) by eight (condition) ANOVA with repeated measures on the last two factors. It was found that, in these trials, there was no significant difference between the three groups ($F(2, 69)<1$, $p=0.942$). Both situation and condition had significant effects on the participants' performance. There was significant situation effect ($F(1, 69)=14.48$, $p=0.000$); participants performed significantly better in the old trials than in the new trials. Condition ($F(7, 483)=47.02$, $p=0.000$) was also significant. Participants performed significantly differently in the eight experimental conditions. Roughly speaking, there were increasing difficulties from condition one to condition eight with the exception that condition one was more difficult than condition two but less difficult than condition three, four, five, six, seven and eight. The interaction between situation and condition was also significant ($F(7, 483)=3.26$, $p=0.002$).

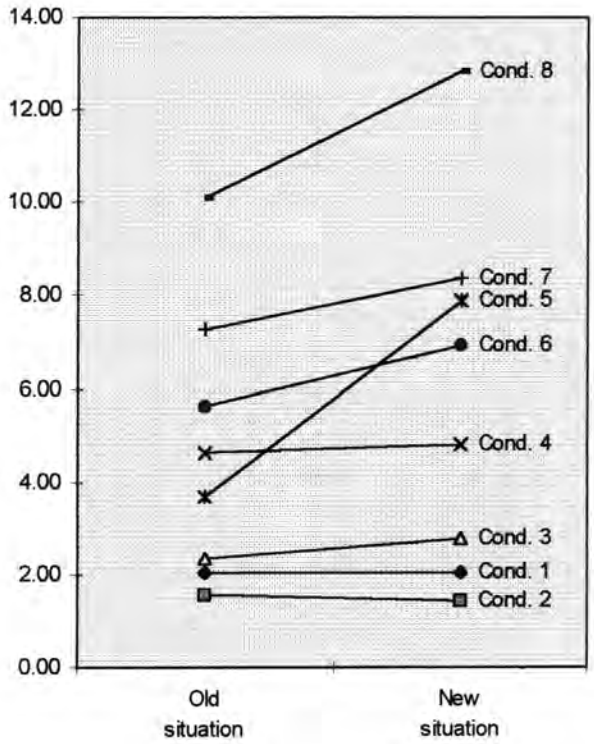
Table 3.2: The mean absolute error in both old and new situations for each participant group and condition in the prediction trials.

	Conditions								Overall Means
	1	2	3	4	5	6	7	8	
Old situation									
Psychiatrist group	2.71	1.58	3.04	3.54	3.75	4.58	6.83	9.21	4.41
Research group	1.71	1.38	2.17	5.42	3.58	6.04	7.58	11.92	4.98
Trainee group	1.71	1.79	1.75	4.92	3.71	6.25	7.42	9.13	4.59
Means	2.04	1.58	2.32	4.63	3.68	5.63	7.28	10.08	4.66
New situation									
Psychiatrist group	1.92	1.08	2.96	5.13	8.29	6.33	9.13	13.13	6.00
Research group	1.67	1.71	2.42	3.67	7.67	7.17	8.63	13.50	5.81
Trainee group	2.46	1.54	2.92	5.58	7.75	7.33	7.33	11.79	5.84
Means	2.01	1.44	2.76	4.79	7.90	6.94	8.36	12.81	5.88
Overall Means	2.03	1.51	2.54	4.71	5.79	6.29	7.82	11.45	5.27

To examine the interaction between situation and condition, the errors in each of the eight conditions were plotted against both situations. Figure one showed the mean absolute errors of all the participants in both situations for each of eight conditions. It could be seen in the graph that errors in the new situation were generally higher than those in the old situation, especially for the high conditions. There was a general trend that the higher the condition the greater the difference between old and new situations, except for condition five. Condition five had the largest differences between old and new situations and this made the significance of the interaction

between condition and situation. Paired sample t-tests were used to compare the absolute errors between the two situations in each of the eight conditions. It was found that the differences reached the significance level of 0.05 only in condition five ($t(71)=4.66$, $p=0.000$) and condition eight ($t(71)=2.02$, $p=0.047$). In both cases, participants performed better in the old situations than in the new situations.

Figure 3.1. Mean absolute errors of all the participants in both situations for the each condition in the prediction trials



The interactions between group and either situation or condition were far from significant. ($F(2, 69)<1$, $p=0.628$ and $F(14, 483)<1$, $p=0.937$ respectively). To analyse whether there existed some subtle differences between groups, errors within each group were further compared. It was found that there was a significant situation effect in the psychiatrist group ($F(1, 23)=13.11$, $p=0.001$), in which mean errors in the new

trials were significantly greater than that in the old trials. However, situation effect was not significant in the research group ($F(1, 23)=2.33$, $p=0.140$). Situation effect in the trainee group was marginally significant ($F(1, 23)=3.54$, $p=0.073$) with the same direction as the psychiatrist group.

Judgement trials

Table 3.3 shows the mean absolute error in both old and new situations for each group and each condition. In these trials, the absolute errors were measured by drug dosage in millilitres not the happiness index as in the prediction trials. The data in this table were analysed using a three (group) by two (situation) by eight (condition) ANOVA with repeated measures on the last two factors. There was no significant effect of group ($F(2, 69)=1.27$, $p=0.288$), but there was a significant effect of condition similar to that in the prediction trials ($F(7, 483)=116.63$, $p=0.000$). The interaction between group and condition failed to reach significance ($F(14, 483)=0.97$, $p=0.480$). The situation effect was not significant ($F(1, 69)<1$, $p=0.420$), neither were the interactions between situation and either group or condition ($F(2, 69)<1$, $p=0.524$; $F(7, 483)=1.25$, $p=0.274$ respectively).

Table 3.3: The mean absolute error in both old and new situations for each participant group and condition in the Judgement trials.

	Conditions								Overall Means
	1	2	3	4	5	6	7	8	
Old situation									
Psychiatrist group	4.96	2.92	4.67	5.79	8.96	11.58	12.67	15.58	8.39
Research group	4.04	4.00	5.75	6.75	9.21	14.25	17.21	20.67	10.23
Trainee group	5.13	3.38	4.08	4.92	8.92	12.33	15.17	19.96	9.23
Means	4.71	3.43	4.83	5.82	9.03	12.72	15.01	18.74	9.29
New situation									
Psychiatrist group	4.58	3.00	4.25	7.50	9.25	8.42	16.54	18.83	9.05
Research group	3.96	2.54	5.29	8.33	11.00	11.29	18.13	19.54	10.01
Trainee group	4.38	4.46	4.75	8.25	8.58	13.13	14.13	18.96	9.58
Means	4.31	3.33	4.76	8.03	9.61	10.94	16.26	19.11	9.55
Overall Means	4.51	3.38	4.80	6.92	9.32	11.83	15.64	18.92	9.42

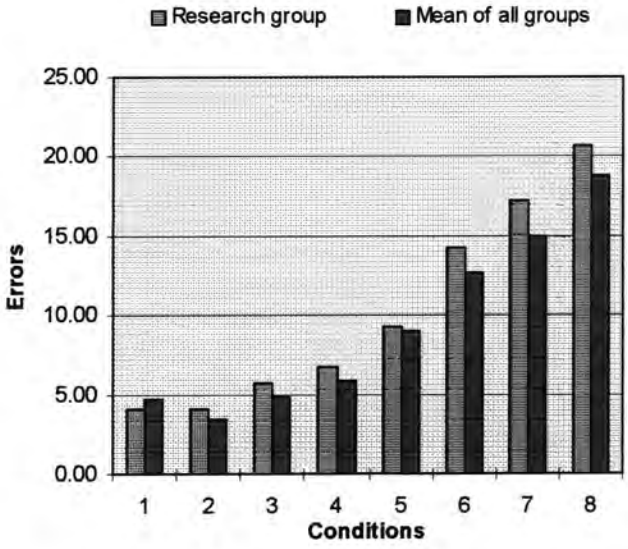
However, by looking at Table 3.3 carefully, it was found that in the old situations, the research group performed worse than the average of the three groups. This happened in seven out of eight conditions, the exception being condition one. The probability of this outcome by chance by Binomial test is 0.035, with a significance level of 0.05. Although, a group effect was not found in the ANOVA with repeated measures by using individual data, it would probably be a Type II error. This could happen when individual variance was relatively large. When comparing groups by their means, individual variance is demolished and the comparison should be clearer. In the

new situation, a similar thing happened, the research group performed worse than the average of the three groups. This happened in six out of eight cases, the exceptions being conditions one and two, and would not reach significance. Alternatively, a claim of "the research group did worse in manic-depressive cases" would reach 0.05 level of significance. The research group did worse than the other two groups in both situations for conditions four to condition eight. The comparison between mean absolute errors of the research group and that of the mean of three groups is showed in Figure 3.2. The figure shows that in most conditions, the research group performed worse than the mean of three groups.

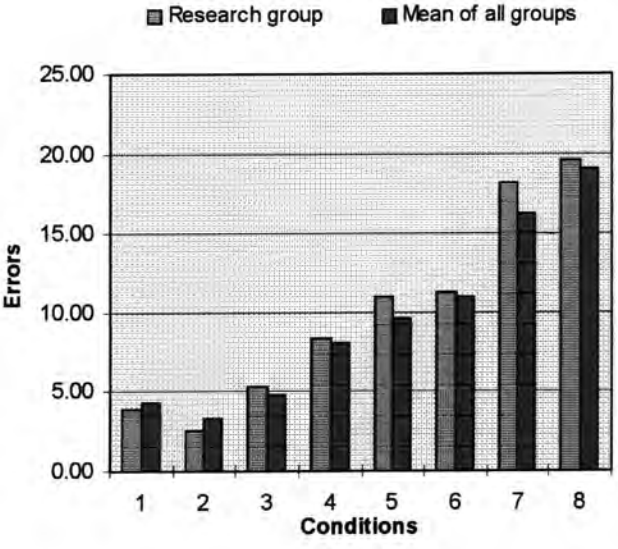
Having discovered the fact that the research group performed worst in most of the eight conditions, absolute errors of the participants were further analysed by grouping the eight conditions into the three mood states, depressive, manics and manic depressive. The inquiry here is to see whether there is a significant interaction between group and mood state, with the research group performing less well with manic depressive cases but with there being no difference between the three groups with depressive cases or manic cases. According to the parameter of the logistic map formulae, the first two conditions were grouped into depressive case, the third condition was classified as a manic case, and the remaining five were grouped into the manic depressive case. The mean absolute errors of the conditions in each category was calculated as the statistical data of each mood state and an ANOVA with repeated measures on mood and situation was used for analysis. It was showed that the interaction between group and mood state did not reach any significance ($F(4, 138)=1.41$, $p=0.235$).

Figure 3.2. Comparison of mean absolute errors between the research group and the group means in both situations of the judgement trials.

Old situations



New situations



Correlations

Table 3.4 showed the correlations between the two observer groups and the psychiatrist group in the prediction trials. The correlations were computed by using each participant's mean absolute error in each situation. The performance of each participant in the psychiatrist group was compared with his/her two matched observers in the research group and the trainee group. The data were analysed with Pearson's correlation. Correlations in the old situation for both comparison pairs were large and highly significant, but correlations in the new situation for both pairs did not reach significance.

Table 3.4: Correlations between the two observer groups and the psychiatrist group in the prediction trials

Correlation	Old trials	New trials	Overall
Research group and Psychiatrist group	R=0.7152, p=0.000	R=0.3137, p=0.136	R=0.5609, p=0.000
Trainee group and Psychiatrist group	R=0.7722, p=0.000	R=0.2773, p=0.190	R=0.5974, p=0.000

Table 3.5 shows the correlations between the two observer groups and the psychiatrist group in the judgement trials. The correlations were computed by using each participant's mean absolute error in each situation. The performance of each participant in the psychiatrist group was compared with his/her two matched observers in both the research group and the trainee group. The data were analysed with

Pearson's correlation. The correlations for the research group in both situations were large and significant, but the correlations for the trainee group were less significant, especially in the new trials, where the correlation did not reach 0.05 significance level.

Table 3.5: Correlations between the two observer groups and the psychiatrist group in the judgement trials

Correlation	Old trials	New trials	Overall
Research group and Psychiatrist group	R=0.6019, p=0.002	R=0.6490, p=0.001	R=0.6157, p=0.000
Trainee group and psychiatrist group	R=0.5713, p=0.004	R=0.3756, p=0.070	R=0.4798, p=0.001

Chaotic behaviour in the prediction trials

The correlation between participants' errors on the prediction task and the parameter of the logistic map formula was high. The same method was used to measure the correlation as in the previous experiment; absolute mean error and the deviation from the optimum A value were compared. The data were analysed with Pearson's correlation.

Table 3.6: The correlations between the parameter Z and mean absolute errors in each group and each situation

Correlation with parameter Z	Psychiatrist group	Research group	Trainee group	Overall
Old trials	0.9525, (8), p=0.000	0.9313, (8), p=0.001	0.9565, (8), p=0.000	0.9577, (8), p=0.000
New trials	0.9433, (8), p=0.000	0.9578, (8), p=0.000	0.9337, (8), p=0.001	0.9521, (8), p=0.000

The correlation of parameter Z with individual participants' absolute errors in each situation were separately also high.

Table 3.7: The correlations between the parameter Z and absolute errors of each individual participant in each situation.

	Old trials	New trials	Overall
Correlation with parameter Z	0.4203, (576), p=0.000	0.5144, (576), p=0.000	0.4666, (1152), p=0.000

Treatment trials (the psychiatrist group)

Similar to the judgement trials, absolute errors in these trials were measured by drug dosage in millilitres. The findings in these treatment trials were similar to that of the previous experiment. There was significant condition effect. ($F(7, 161)=33.85$, $p=0.000$). In addition, there was significant order effect ($F(1, 23)=5.80$, $p=0.024$), participants performed significant better in the last four trials than the first four. However, there was no significant interaction between these two effects ($F(7, 161)<1$, $p=0.457$).

Questionnaire

Table 3.8 showed the number of participants who answered correctly in each group for each question. There were overall 24 participants in each group.

Table 3.8: Number of participants answered correctly out of 24 participants in each group for each question

	Question 1	Question 2	Question 3	Question 4	Total
Psychiatrist group	23	10	6	3	42
Research group	21	16	9	4	50
Trainee group	22	15	7	2	46

The data was analysed by a one sample Chi-square on each question. For the first three questions, numbers of correct responses were used, while incorrect responses were used for question four to avoid having expected frequencies of less than five. It was found that there was no significant group effect for any of the four questions.(χ^2=0.0909, p=0.9556; χ^2=1.5122, p=0.4695; χ^2=0.6364, p=0.7275 and χ^2=0.0925, p=0.9535 respectively)

Discussion

In the prediction trials, no significant difference between groups was observed. The research group participants performed similarly to their counterparts in the psychiatrist group. The finding of an insignificant group effect between the psychiatrist group and the research group is consistent with the last experiment, in which each participant was only required to respond to five of the eight condition in each situation. In the present experiment, participants were required to answer all of the eight conditions in each situation. The additional experimental group, the trainee group, in this experiment also performed similarly to the other two groups. As mentioned in the experimental design section, while the psychiatrist group represented an executive group with a problem-solving goal and the research group represented a group of observers with a rule-searching goal, the trainee group were observers with a problem-solving goal. The insignificant group effect between the three groups implies that neither role nor goal imposed any significant group differences on the prediction trials.

The result also showed that there existed a significant situation effect in general. Participants performed significantly better in old situations than in new situations. This may be simply explained by the effect of memory. Participants had memories of the cases in the old situations, and this enhanced their performance, while memory could not be used in the new situation. By further analysing the differences between old and new situations, it was found that the difference was significant in the psychiatrist group, marginally significant in the trainee group but not significant in the research group. It is possible to say that the behaviour of the trainee group lay between

those of the psychiatrist group and the research group. A possible interpretation is that the differences between old and new situations are a matter of learning type. Learning can be done implicitly or explicitly. If the difference between old and new situations is large, participants have learnt implicitly and stored in memory the learning instances. If there is no difference between old and new situations, participants have learnt explicitly and they have learned rules as well as instances. According to this view, the difference in behaviour across these three groups may be explained by the type of learning they had employed. That is the psychiatrist group tended to employ the strategy of implicit learning. The research group tended to employ the strategy of explicit learning. The trainee group lay between these two extremes, showing some evidence of rule learning. The reason the trainee group lay between the other two is not quite clear. It may be possible that both role and goal had imposed an effect on learning.

The interaction between situation and condition was unexpected. Further analysis of the situation effect in each of the eight conditions showed that only in condition five and condition eight did the situation effect reach the significance level of 0.05. It is possible that the case in condition eight is a Type I error, since it is one case out of eight cases and with $p=0.047$ it is just above the significance level of 0.05. Thus, it seems that the most unexpected result is the significance of condition five. Two explanations may be possible. First, from condition three to condition eight, the differences between old and new situations have the same trend, that is in all these cases, participants performed better in the old situations than in the new situations. The only problem is whether the differences reach significance level and consequently some conditions reach significance and some do not. It is a matter of chance that in the

present experiment condition five can reach significance while the others did not. Another explanation is that strong implicit learning has occurred in condition five and thus it showed a large difference between old and new situations while the overall performance in condition five still lay between those in condition four and condition six. However, the exact reason of stronger implicit learning is unclear. Perhaps, the chaotic behaviour in condition five is more impressive for memory. As far as the present finding is concerned, no strong claim can be made on that point.

The situation effect was significant in the prediction trials; it is reasonable to think that memory of the old trials makes the difference between the old situations and the new situations. However the results showed that situation in the judgement trials did not reached significance. Why is there a difference between the prediction trials and the judgement trials? If memory of old trials makes participants perform better in the old cases, why is it so in the prediction trials but not in the judgement trials? The phenomenon probably showed that participants' memory was purposeful and selective. Participants tended to memorise what particular input would produce what particular output in order to grasp the behaviour of the system. This might be their major concern and this was what they were asked in the prediction trials. On the other hand, it seems that they had low motivation to remember what they had input in any particular cases, since their past decisions would not help them to understand the system. This would not be their major concern and this was what they were asked in the judgement trials. Therefore, memory played a more important role in the prediction trials than in the judgement trials, and that is why there is significant situation effect in the prediction trials but not in the judgement trials.

It is not unexpected that the research group performed worse than the other two groups in the judgement trials, although the ANOVA showed no statistical difference. Basically, the goal of the psychiatrist group and the trainee group was to learn how to make a judgement based on information about inputs and the outcomes. In the judgement trials, both of these two groups performed similarly. It can be probably said that there is no effect of role in the judgement trials. No matter whether the participants were executors or observers, they performed similarly in the judgement trials as long as their goal was the same. The choice of goal determines whether participants engages in problem solving or rule searching. The goal of the research group did not lead them to focus on judgement. That is why they seemed to perform worse in the judgement trials. Thus it appears that in the present control task, to know the rule and to do the judgement are two different things. This finding seems to be consistent with Berry and Broadbent's (1984) experiment. This experiment showed that verbal instruction significantly improved the ability to answer questions but had no effect on control performance. Even when the participants were explicitly told the rules behind the system, they could not perform better, although they knew the rule of the system better. A duality of rule knowledge and performance seems to exist. If so, in the present experiment, it is possible that by concentrating on rule learning, the research group participants did less well on performance. However, this discussion needs to be treated with caution, because the difference between groups was not statistically significant.

Correlation of intergroup performance is another aspect worth exploring. In the judgement trials, the correlation between the psychiatrist group and the two observer

groups is R^2=0.3791 for the research group and R^2=0.2302 for the trainee group. The higher correlation between the psychiatrist group and the research group than that between the psychiatrist group and the trainee group seemed to further show that the trainee group learnt better than the research group. The correlation here is interpreted as that the lower the correlation is the more is the participants internalise the observations and make their own judgements. It means that if a participant is not able to judge, he/she will tend to follow the executor's judgement based on their memory of the treatment trials. On the contrary, if the participant has his own idea of the way of judgement, he/she will not follow the executor's judgement so much. In other words, the more a group has learnt, the less the correlation with the psychiatrist group should be. This difference of learning between the research group and the trainee group is thus reflected in the correlation mentioned above. The comparison of the two correlations above was further investigated in each of the two situations. In the old trials, correlations between the psychiatrist group and both the research group and the trainee group is similar, with R^2=0.3623 for the research group and R^2=0.3624 for the trainee group. However, in the new trials, the difference between these two correlations is apparent, with R^2=0.4212 for the research group and R^2=0.1411 for the trainee group. This result further shows that the trainee group was able to make more independent judgements in the new situations than the research group. This result is consistent with the findings of their performance in the judgement trials that the research group performed worse than the trainee group, and it is because the research group learnt worse than the trainee group. In the prediction trials, the correlation between the psychiatrist group and both the research group and the trainee group are similar. The correlations is R^2=0.3146 for the research group and R^2=0.3569 for the trainee group

and both reached significance. Furthermore, both groups have a significant correlation with the psychiatrist group in the old situations but not in the new situations. Thus, it seems that the manipulation of goals has no effect on these trials. The goal have only manipulated a effect on the judgement in the judgement trials.

The correlations between the mean absolute errors and the parameter of the logistic map formula were very high. There was a overall correlation of R^2=0.9172 in the old situations and of R^2=0.9065 in the new situations. There seemed to be no significant difference between old and new situations. Similarly, there was no observable difference in the correlations between the three experimental groups. This correlation remained high as in the last experiment. These high correlations mean there is a high predictability of group behaviour of forecasting in such a chaotic situation. However, when all data points were used in the correlation with Z, the correlation drops to R^2=0.1767 for old situations and R^2=0.2646 for new situations. However, this does not happen in the group comparison, only in the individual comparison. It may be possible that there are too few data points in group correlations so the analysis is easily affected by random factors. The analysis in the present experiment is original and further analysis in this area is worthwhile when the application of chaos theory in the social sciences is becoming more significant.

To summarise, in the prediction trials, the insignificant group differences do not show a strictly significant goal effect, while the significant situation effect probably shows the existence of implicit learning. The differences of the significance of situation across the three experimental groups may imply that implicit learning occurs to the

greatest extent in the psychiatrist group and least in the research group. The reason for the interaction between situation and condition is not quite clear. There was no significant situation effect in the judgement trials as in the prediction trials; this may indicate that implicit learning is selective and purposeful. The fact the research group performed worst in the judgement trials probably shows that the group concentrated on the rule-searching, which undermined its ability on judgement. The intergroup correlations on performance further show that the trainee group learned better than the research group while the psychiatrist group learned similarly to the trainee group. Finally, this experiment further consolidated the correlation between participants' absolute errors and the derivation of A.

Chapter Four

General Discussion

GENERAL DISCUSSION

The findings of the second experiment were basically consistent with the findings in the first experiment. In the prediction trials of both experiments, there was no significant difference between the research group and the psychiatrist group while situation effect was significant in the psychiatrist group but not in the research group. The participants' absolute errors systematically varied with the parameter of the logistic map formula in both experiments. However, there were also some inconsistencies between the two experiments. In the first experiment, there were significant group differences in particular conditions, but this did not happen in the second experiment. Furthermore, although not significant, a group difference was observed in the judgement trials of the second experiment, but this was not observed in the first experiment. These inconsistencies may be explained by the difference of the experimental design between the two experiments.

In the first experiment, prediction trials and judgement trials were randomly selected in five out of the eight conditions to each participant and thus each participant received a different combination of conditions. However, in the second experiment, this variable was more carefully controlled; trials were given from each of the eight conditions to the participants and participants across groups received identical trials with the same order. Therefore, there was no difference in difficulties of the trials across groups and it was believed that comparisons across groups should have been more reliable. Also, the second experiment should have provided a more precise

comparison between groups and situations. This may explain why group differences happened in particular conditions in the first experiment but not in the second experiment. In the second experiment, no significant group differences were observed in any individual condition. The group differences in particular conditions in the first experiment might possibly be Type I errors.

A non-significant group difference was observed in the judgement trials of the second experiment, but this was not observed in the first experiment. It is believed that the failure to find a group difference in the first experiment was a Type II error and with better control of the variables in the second experiment the Type II error was reduced. Because the second experiment was more reliable than the first, the following discussion will be mainly based on the second experiment.

In comparing the present experiment with Geddes' experiments (Geddes and Stevenson 1997), some differences as well as some similarities were found. In Geddes' experiments, the rule searching groups performed significantly better than the problem solving groups in the prediction trials. The difference between old and new situations was significant in the problem solving groups but not in the rule searching groups. In addition, the rule searching groups did not perform significantly worse than the problem solving groups in the test trials. In the present experiment, there was no significant difference between groups in the prediction trials. The difference between old and new situations was significant in the problem solving groups but not in the rule searching group, while the rule searching group performed worse than the problem solving groups in the judgement trials.

In the Clegg control task a definite input can be obtained by considering the previous mood state of Clegg. However, in the task of the present experiment, a definite input is not possible by only considering the previous state of the system; the output will remain oscillating until a stable input has been maintained for several successions. In the Clegg control task, there are only twelve possible states in the system; while in the task of the present experiment, any value between zero and fifty could arise in the system. In the Clegg control task, the same rule applies for different states of the system; while in the task of the present experiment, the rules for different states of the system are different.

It is believed that the reason for the differences between the two sets of studies is the difference in the complexity of the control tasks in the two experimental series. In the Clegg control task examined by Geddes, it was possible for rule searching participants to understand the rules exhaustively within thirty trials, while this may not be the case in the present experiment. If participants understood the rules exhaustively, they would be able to apply the rules to judgement and prediction accurately. However, before that they cannot perform better than the other participants. This is the case in the present experiment; the rules were too complicated for participants to understand exhaustively within the limited trials in the experiment. Before they understood them completely, how could they make full use of them for prediction and judgement? The other two groups had already spent their effort in implicit learning through practice; while most of the effort of the research group was to understand the rules but this group had not got any benefit in the ability to make judgements in the present experiment by examining the rules. That may be a reasonable account of the

fact that the research group performed worse in the judgement trials than the other two groups, in contrast to Geddes' results.

However, there is also an important similarity between the findings of the two series of experiments. Both series of experiments showed that learning was affected by goals. For a problem solving goal, participants would tend to learn implicitly and their learning would rely more on memory for instances. For a rule searching goal, participants would tend to learn explicitly and their learning would rely more on identification of rules. This was indicated in the findings of the prediction trials in both experimental series, the problem solving groups performed significantly better in the old situations than in the new situations while there were no significance differences between the two situations for the rule searching groups. The preference of implicit learning or explicit learning in individual groups was observed, although the difference between groups was reduced in the present experiment, for the reasons outlined above, compared to Geddes' experiments.

From the findings of the present experiment, it is still not clear whether the research group had attained a better understanding of the underlying rules of the system. The performance of the research group in the questionnaire section was not significantly better than the other two groups. Perhaps more questions could be designed to make clear any possible differences between the groups. It is important to notice also that unlike Geddes' study, forecasting in the present study was in a situation of uncertainty. Although in Geddes' experiment, noise was introduced into the rule to alter the output so that the rule was not too obvious, participants who

grasped the rule knew exactly the whole mechanism of the system and the extent to which the noise affected it. This was not the case in the present experiment. In everyday life, prediction seems most needed in situations where the rules are complex and difficult to grasp (e.g. economic situations). Unless the rules are known exhaustively or almost exhaustively, explicit learning seems to be unable to improve the performance of forecasting. On the contrary, under uncertainty, implicit learning seems to be more important to the quality of performance. This suggests that practice and experience play important roles in learning before rules can be explicitly and accurately known. If this interpretation is true, rule searching will be better for some situations, while learning through practice will be better for other situations. For those activities where the behaviour can be described exhaustively by rules (e.g. using computer software), it is better to employ rule learning. For those activities that cannot be easily described by rules or any uncertain situations, implicit learning may be more effective for forecasting. To verify this point of view, further studies could be done on control tasks of different natures or levels of difficulty. The use of a more extensive training period could also be investigated to see if longer training might enhance rule learning by the research group participants. In these ways, it could be also shown whether the findings of the present experiment could be extended to a wider area of forecasting.

Regarding chaos theory, it might be interesting to look further into the issues involved. Nevertheless, the focus of investigation in the present study is the effects of goals on forecasting, rather than to investigate chaos theory in general. The chaos control task is only acting as a tool for the investigation of this topic. The reason for

using a chaos control task in this study is because it is the most convenient method that can be thought of to produce some unpredictable data for assessing the forecasting ability of participants. More importantly, Harvey and Bruce have used such control tasks in their experiments and it is easier to make comparisons if a control task of similar nature is used in the present experiment.

Look back to the contemporary research in forecasting. Most research emphasises the objective methods of forecasting and their applications. They work out statistical models for forecasting which involve various types of mathematical calculation to examine the trend of phenomena and make predictions about the future. Many of these modelling methods and their applications are investigated in particular areas. Research is done in the areas from natural sciences to social sciences. The themes range from rainfall forecast in some geographic regions (Mason 1998) to unemployment rate forecast in a particular country (Montgomery, Zarnowitz, Tsay and Tiao 1998). However, this line of research involves the models for or the methods of forecasting without considering any particular cognitive factors.

The closest area to the study of cognitive factors of forecasting is the psychology of decision making in cognitive psychology. It concerns subjective utility and the values of the person making the judgement. It makes comparisons between the subjective perceived value and the objective expected value of the outcome. The focus of investigation is the inconsistency between the subjective perception of some outcome values and the actual expected value obtained by considering the probability of the outcomes. The research shows that human decision making consists of heuristics

and various kinds of biases in judgement. However, existing decision making theories in cognitive psychology have not been extended to the study of forecasting. Existing theories of decision making mainly examine present events. Although, occasionally, some future events are mentioned, the events have a known probability. Real forecasting involves future events in which the probability of the outcome is uncertain.

Most of the few studies on psychology of forecasting rest on the research on time-series analysis. A list of figures are presented to participants and their forecasting behaviours are observed. For example, in one study (Harvey, Bolger & McClelland 1994), a series of criminal rate record in an underground railway system was employed. Participants were given successive data points of the criminal rate and were asked to predict the next figures. The intention of this line of investigations is to observe how people make predictions in time-series situations. However, this line of studies did not involve any manipulation of factors nor learning, so it is not the interest of the present research. Although both mathematical models for forecasting and time-series forecasting behaviour studies provide good foundations for studying forecasting, little concern is given to the study of how to improve forecasting and by which method, as well as the factors affecting improvement.

Those research on mathematical models provide rational models of forecasting and a normative view of forecasting (the ideal way of forecasting) and the time-series forecasting studies about human decision making and judgement provide a descriptive view of human forecasting (the actual way of forecasting). However, to suggest a way to improve human forecasting, a prescriptive model of learning to forecast is needed.

Theoretical models can provide accurate predictions only in ideal situations. In real situations, revisions have to be made according the actual circumstances. Only the grasp of good judgmental forecasting by a decision maker can bridge the gap between prediction by a theoretical model and the actual trend of a real situation. Also, only by better understanding of the cognitive factors affecting judgmental forecasting, can a prescriptive model of learning be derived. The use of a control task in the present research to investigate the cognitive factors affecting judgmental forecasting will be a worthy trial in the area of forecasting. In future, if more research could be done on the cognitive factors affecting forecasting, there will be better understanding of ways to improve judgmental forecasting.

Appendix One

Raw data and Additional statistics for Experiment One

APPENDIX ONE

Table of Contents:

Appendix 1.7: Paired samples t-test analyses of the situation difference in each condition and each group of the prediction trials by comparing the absolute errors in Appendix 1.2.

Appendix 1.8: Independent samples t-test analyses of the group difference in each condition and each situation of the prediction trials by comparing the absolute error in Appendix 1.2.

Appendix 1.9: Pearson's Correlation between mean absolute errors and Z, the derivation of parameter A in the logistic map formula.

Appendix 1.10: Signed errors of the participants in the judgement trials

Appendix 1.11: Absolute errors of the participants in the judgement trials

Appendix 1.12: A mixed ANOVA of the judgement trials for the overall group and condition effects

Appendix 1.13: One-way ANOVA for each condition in the judgement trials

Appendix 1.14: Mean absolute errors of the 30 psychiatrist group participants in the treatment trials

Appendix 1.15: A simple ANOVA with repeated measures of the treatment trials

Appendix 1.16: Pearson's Correlation between mean absolute errors in the both prediction trials and judgement trials for both experimental group.

Appendix 1.1: Signed errors of the participants in the prediction trials

Psychiatrist Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
1	4	-11			-2	-2		8	0			-6	-6		10	-27
2	1	-1				-1	-5	-4	-5	-1		-1	-2		-5	
3	2		0	3		0	-3		2			-2	-17	-11		-1
4		0	-16	-33	21			8	5	-1	10			18		40
5	0		0	3	6		-6				1	-2	-2	-16		28
6		-1	-2	-5	10		-6		-1	0		-4	-11	0		
7	0	-1		-18	0		-1		2		-1			-15	-3	-10
8	0	0				-5	0	-12	1			-8	-8	5		-10
9		6	-11			-11	-10	-16		6	-4	5		-5		-2
10	0	0	0		1		0			1	0			-8	-22	-7
11	0		-6	-1	2			4	-3		-1		5	-15		4
12	0	0	8		-2	-1			1	1			-4	-19	-29	
13	-1	-3		0		-2		0		-2	4	1	9			-6
14	8		1	-1	-1		2		0	-2	-8		-7			-4
15	5	7		-8	-1	12			-11	7		-1		-9	-2	
16	-3	4				-3	-2	-6		3	-1	3	0		10	
17			0	-1	1	1		5	-2	-1	1			-7	0	
18	1	9	6		-3		36		11		2	-1	-1		-9	
19			-3		4	-4	-4	29			-1		6	-3	8	1
20			7	0	-4		-3	27	-9	-4	10	8			0	
21	1		-5		9	4		38			-6	17		0	-2	23
22	0			-1	-10		-1	6	-1			-7	0	-6		-7
23		7		4	10		5	9		11		-3	-11	15		-13
24	5	4				-19	18	-10	9				-22	-3	-8	28
25			-2		-1	1	-4	-14		2		-12		-5	-30	1
26		-1			0	1	-1	6	1	1				-7	-9	-8
27	0	0	-2		5		0		1	2	-2		-13		-16	
28		1	-2	-2		-2	0		2			-9	-10	4	-5	
29	1	3	-3	2	5				2	4		2			-18	-23
30	0	1		-4	0	0			0		4	3	-4		-2	
	1.20	1.20	-1.67	-3.88	2.27	-1.82	0.75	4.59	0.24	1.59	0.50	-0.89	-5.16	-4.35	-6.95	0.37

Research Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
31	4		-14	-6		0		-12	-11	5	-1		2		-10	
32	0				0	-3	2	3		2	0			7	-16	-17
33	0	1		0	0	3						-4	6	3	-7	-17
34	-1	0		-7	11	-20			1	1		3		-10		12
35			1	-3	0		-2	-14	7	5	-2	3				-7
36				-10	-10	13	-8	-15		6	-2	-4	-7	0		
37	1	0	0		-9	15				0	-2	-8	5			-6
38		1	-1	0		3		-6	-1		1	-9	-5			-8
39	-1	2				5	1	-3	1	-1	-1	14			-23	
40		-1	0	1		-11	2			-2	2	-5			-12	-4
41		0	-4		-10		-13	42	5	1			-11	8	-26	
42	1	0		5	-14	-12			0			-7	-9		-14	12
43		-2	5	0		-5	0		1	0	0	-5		-3		
44	7	-1			-2	1		0	0		0			0	0	-3
45	4	2		1			19	-15		-3			5	19	-14	-14
46	-1				-13	-1	-8	-14		4	-3			-8	-3	9
47		1		-2	1	1	-1			-1			-3	-8	-14	6
48	0	1			-1	22		-2			-2	-5	-6	3	-9	
49	0			-1	5		-20	6	-2	1	0	-2				-10
50		0		-4		-4	3	-22	6	0	-2				-18	-19
51	-2	0				-3	-24	-14	2	0	1	-4			-8	
52	2			-2		-1	-5	6		0		3	1	-8	-1	
53		0	38	-13		32	0		4	0		7	15	3		
54		0		6	18	-14	22			0	0			-9	-21	
55			1	1	1	1		-16		0	-2	-1		-3	-2	
56		0		1		0	1	9	1	4			-5		6	6
57		1		-3		-3	-1	-1				-2	-3	-4	-13	-10
58	0	1	1				-1	1	1	0			-3	9		20
59				0	1	-8	5	11	3		1	-2	-7		-12	
60		0	3		3	-5	7		5	2			5	2	21	
	0.93	0.29	2.73	-1.80	-1.12	0.24	-1.00	-2.80	1.35	1.04	-0.67	-1.56	-1.18	0.06	-9.33	-2.94

Note: In the table, positive figures mean over estimation; negative figures mean under estimation; zeros mean participants have estimated the figures correctly.

Appendix 1.2: Absolute errors of the participants in the prediction trials

Psychiatrist Group

	Old Situation								New Situation							
	Condition								Condition							
Subject No.:	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	4	11			2	2		8	0			6	6		10	27
2	1	1				1	5	4	5	1		1	2		5	
3	2		0	3		0	3		2			2	17	11		1
4		0	16	33	21			8	5	1	10			18		40
5	0		0	3	6		6				1	2	2	16		28
6		1	2	5	10		6		1	0		4	11	0		
7	0	1		18	0		1		2		1			15	3	10
8	0	0				5	0	12	1			8	8	5		10
9		6	11			11	10	16		6	4	5		5		2
10	0	0	0		1		0			1	0			8	22	7
11	0		6	1	2			4	3		1		5	15		4
12	0	0	8		2	1			1	1			4	19	29	
13	1	3		0		2		0		2	4	1	9			6
14	8		1	1	1		2		0	2	8		7			4
15	5	7		8	1	12			11	7		1		9	2	
16	3	4				3	2	6		3	1	3	0		10	
17			0	1	1	1		5	2	1	1			7	0	
18	1	9	6		3		36		11		2	1	1		9	
19			3		4	4	4	29			1		6	3	8	1
20			7	0	4		3	27	9	4	10	8			0	
21	1		5		9	4		38			6	17		0	2	23
22	0			1	10		1	6	1			7	0	6		7
23		7		4	10		5	9		11		3	11	15		13
24	5	4				19	18	10	9				22	3	8	28
25			2		1	1	4	14		2		12		5	30	1
26		1			0	1	1	6	1	1				7	9	8
27	0	0	2		5		0		1	2	2		13		16	
28		1	2	2		2	0		2			9	10	4	5	
29	1	3	3	2	5				2	4		2			18	23
30	0	1		4	0	0			0		4	3	4		2	
	1.60	3.00	4.11	5.38	4.45	4.06	5.35	11.88	3.29	2.88	3.50	5.00	7.26	8.55	9.89	12.79

Research Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
31	4		14	6		0		12	11	5	1		2		10	
32	0				0	3	2	3		2	0			7	16	17
33	0	1		0	0	3						4	6	3	7	17
34	1	0		7	11	20			1	1		3		10		12
35			1	3	0		2	14	7	5	2	3				7
36				10	10	13	8	15		6	2	4	7	0		
37	1	0	0		9	15				0	2	8	5			6
38		1	1	0		3		6	1		1	9	5			8
39	1	2				5	1	3	1	1	1	14			23	
40		1	0	1		11	2			2	2	5			12	4
41		0	4		10		13	42	5	1			11	8	26	
42	1	0		5	14	12			0			7	9		14	12
43		2	5	0		5	0		1	0	0	5		3		
44	7	1			2	1		0	0		0			0	0	3
45	4	2		1			19	15		3			5	19	14	14
46	1				13	1	8	14		4	3			8	3	9
47		1		2	1	1	1			1			3	8	14	6
48	0	1			1	22		2			2	5	6	3	9	
49	0			1	5		20	6	2	1	0	2				10
50		0		4		4	3	22	6	0	2				18	19
51	2	0				3	24	14	2	0	1	4			8	
52	2			2		1	5	6		0		3	1	8	1	
53		0	38	13		32	0		4	0		7	15	3		
54		0		6	18	14	22			0	0			9	21	
55			1	1	1	1		16		0	2	1		3	2	
56		0		1		0	1	9	1	4			5		6	6
57		1		3		3	1	1				2	3	4	13	10
58	0	1	1				1	1	1	0			3	9		20
59				0	1	8	5	11	3		1	2	7		12	
60		0	3		3	5	7		5	2			5	2	21	
	1.60	0.67	6.18	3.30	5.82	7.44	6.90	10.60	3.00	1.65	1.22	4.89	5.76	5.94	11.90	10.59

Appendix 1.3: The mean absolute errors for each participant in each situation calculated from Appendix 1.2.

Psychiatrist Group			Research Group		
Subject No.:	Old Situation	New Situation	Subject No.:	Old Situation	New Situation
1	5.40	9.80	31	7.20	5.80
2	2.40	2.80	32	1.60	8.40
3	1.60	6.60	33	0.80	7.40
4	15.60	14.80	34	7.80	5.40
5	3.00	9.80	35	4.00	4.80
6	4.80	3.20	36	11.20	3.80
7	4.00	6.20	37	5.00	4.20
8	3.40	6.40	38	2.20	4.80
9	10.80	4.40	39	2.40	8.00
10	0.20	7.60	40	3.00	5.00
11	2.60	5.60	41	13.80	10.20
12	2.20	10.80	42	6.40	8.40
13	1.20	4.40	43	2.40	1.80
14	2.60	4.20	44	2.20	0.60
15	6.60	6.00	45	8.20	11.00
16	3.60	3.40	46	7.40	5.40
17	1.60	2.20	47	1.20	6.40
18	11.00	4.80	48	5.20	5.00
19	8.80	3.80	49	6.40	3.00
20	8.20	6.20	50	6.60	9.00
21	11.40	9.60	51	8.60	3.00
22	3.60	4.20	52	3.20	2.60
23	7.00	10.60	53	16.60	5.80
24	11.20	14.00	54	12.00	7.50
25	4.40	10.00	55	4.00	1.60
26	1.80	5.20	56	2.20	4.40
27	1.40	6.80	57	1.80	6.40
28	1.40	6.00	58	0.80	6.60
29	2.80	9.80	59	5.00	5.00
30	1.00	2.60	60	3.60	7.00
Mean	4.85	6.73	Mean	5.43	5.61

Appendix 1.4: A simple mixed ANOVA for analysing the interaction between situation and group in the prediction trials by comparing the mean absolute error of each participant in each situation for each experimental group

```
* * * * * * Analysis of Variance * * * * * *

        60 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         2 non-empty cells.

         1 design will be processed.
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL          932.70      58     16.08
GROUP                      2.21       1      2.21       .14      .712
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'SITUATION' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL          473.88      58      8.17
SITUATION                 31.72       1     31.72      3.88      .054
GROUP BY SITUATION        21.42       1     21.42      2.62      .111
```

Appendix 1.5: Independent t-test analyses of the group difference in the prediction trials by comparing the mean absolute errors of each participant in each situation in Appendix 1.3.

Analyses for old situation:

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
ERROR IN OLD SITUATION				
PSYCHIATRIST	30	4.8533	3.929	.717
RESEARCH	30	5.4267	3.991	.729

Mean Difference = -.5733

Levene's Test for Equality of Variances: F= .002 P= .966

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	-.56	58	.577	1.023	(-2.621, 1.474)
Unequal	-.56	57.99	.577	1.023	(-2.621, 1.474)

Analyses for new situation:

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
ERROR IN NEW SITUATION				
PSYCHIATRIST	30	6.7267	3.298	.602
RESEARCH	30	5.6100	2.502	.457

Mean Difference = 1.1167

Levene's Test for Equality of Variances: F= 2.349 P= .131

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	1.48	58	.145	.756	(-.397, 2.630)
Unequal	1.48	54.08	.145	.756	(-.399, 2.632)

Appendix 1.6: Paired sample t-test analyses of the situation difference in the prediction trials by comparing the mean absolute errors of each participant in each group in Appendix 1.3.

- - - t-tests for paired samples - - -

Psychiatrist group

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	30	.441	.015	6.7267	3.298	.602
OLD				4.8533	3.929	.717

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
1.8733	3.857	.704	2.66	29	.013
95% CI (.433, 3.314)					

- - - t-tests for paired samples - - -

Research group

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	30	.219	.244	5.6100	2.502	.457
OLD				5.4267	3.991	.729

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
.1833	4.220	.770	.24	29	.814
95% CI (-1.393, 1.759)					

Appendix 1.7: Paired samples t-test analyses of the situation difference in each condition and each group of the prediction trials by comparing the absolute errors in Appendix 1.2.

Psychiatrist Group

Condition 1

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	15	.246	.376	3.2667	3.900	1.007
OLD				1.8000	2.513	.649

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
1.4667	4.086	1.055	1.39	14	.186
95% CI (-.797, 3.730)					

Condition 2

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	13	.900	.000	3.0769	3.174	.880
OLD				2.5385	2.696	.748

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
.5385	1.391	.386	1.40	12	.188
95% CI (-.303, 1.379)					

Condition 3

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	12	.590	.044	3.8333	3.713	1.072
OLD				4.7500	4.920	1.420

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
-.9167	4.055	1.171	-.78	11	.450
95% CI (-3.494, 1.661)					

Condition 4

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	11	-.436	.180	3.8182	2.857	.861
OLD				2.9091	2.343	.707

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
.9091	4.415	1.331	.68	10	.510
95% CI (-2.058, 3.876)					

Condition 5

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	12	.243	.446	5.8333	4.108	1.186
OLD				4.5833	3.655	1.055

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
1.2500	4.789	1.382	.90	11	.385
95% CI (-1.793, 4.293)					

Condition 6

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	12	-.295	.351	6.5000	4.890	1.412
OLD				5.0833	5.885	1.699

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
1.4167	8.691	2.509	.56	11	.584
95% CI (-4.107, 6.940)					

Condition 7

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	12	-.095	.769	10.4167	8.458	2.442
OLD				6.1667	10.599	3.060

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
4.2500	14.175	4.092	1.04	11	.321
95% CI (-4.759, 13.259)					

Condition 8

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	13	.010	.974	13.0769	12.446	3.452
OLD				12.3077	10.443	2.896

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
.7692	16.167	4.484	.17	12	.867
95% CI (-9.003, 10.541)					

Research Group

Condition 1

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	8	.231	.581	2.2500	3.615	1.278
OLD				2.0000	2.390	.845

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
.2500	3.845	1.359	.18	7	.859
95% CI (-2.966, 3.466)					

Condition 2

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	15	.138	.625	1.0000	1.254	.324
OLD				.6000	.828	.214

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
.4000	1.404	.363	1.10	14	.288
95% CI (-.378, 1.178)					

Condition 3

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	7	-.518	.233	1.4286	.787	.297
OLD				3.1429	5.080	1.920

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
-1.7143	5.529	2.090	-.82	6	.443
95% CI (-6.829, 3.401)					

Condition 4

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	14	.235	.419	4.0714	2.303	.615
OLD				3.2857	4.084	1.092

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
.7857	4.191	1.120	.70	13	.495
95% CI (-1.635, 3.206)					

Condition 5

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	9	.613	.079	6.5556	2.351	.784
OLD				5.4444	5.270	1.757

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
1.1111	4.256	1.419	.78	8	.456
95% CI (-2.161, 4.383)					

Condition 6

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	15	-.022	.939	4.7333	3.283	.848
OLD				8.3333	9.686	2.501

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
-3.6000	10.294	2.658	-1.35	14	.197
95% CI (-9.302, 2.102)					

Condition 7

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	15	.097	.732	13.8667	7.279	1.879
OLD				7.6000	8.052	2.079

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
6.2667	10.320	2.665	2.35	14	.034
95% CI (.550, 11.983)					

Condition 8

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	11	.194	.567	11.1818	5.564	1.678
OLD				8.2727	7.157	2.158

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
2.9091	8.166	2.462	1.18	10	.265
95% CI (-2.579, 8.397)					

Appendix 1.8: Independent samples t-test analyses of the group difference in each condition and each situation of the prediction trials by comparing the absolute error in Appendix 1.2.

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 1				
PSYCHIATRIST	20	1.6000	2.257	.505
RESEARCH	15	1.6000	1.993	.515

Mean Difference = .0000

Levene's Test for Equality of Variances: F= .355 P= .556

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	.00	33	1.000	.734	(-1.494, 1.494)
Unequal	.00	32.04	1.000	.721	(-1.468, 1.468)

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 2				
PSYCHIATRIST	20	3.0000	3.340	.747
RESEARCH	21	.6667	.730	.159

Mean Difference = 2.3333

Levene's Test for Equality of Variances: F= 24.908 P= .000

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	3.13	39	.003	.747	(.823, 3.844)
Unequal	3.06	20.73	.006	.764	(.745, 3.922)

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 3				
PSYCHIATRIST	18	4.1111	4.324	1.019
RESEARCH	11	6.1818	11.286	3.403

Mean Difference = -2.0707

Levene's Test for Equality of Variances: F= 3.341 P= .079

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	-.70	27	.487	2.938	(-8.101, 3.959)
Unequal	-.58	11.82	.571	3.552	(-9.812, 5.671)

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 4				
PSYCHIATRIST	16	5.3750	8.570	2.143
RESEARCH	20	3.3000	3.585	.802

Mean Difference = 2.0750

Levene's Test for Equality of Variances: F= 2.730 P= .108

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	.98	34	.332	2.110	(-2.215, 6.365)
Unequal	.91	19.20	.376	2.288	(-2.714, 6.864)

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 5				
PSYCHIATRIST	22	4.4545	5.021	1.071
RESEARCH	17	5.8235	5.897	1.430

Mean Difference = -1.3690

Levene's Test for Equality of Variances: F= 2.538 P= .120

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	-.78	37	.439	1.749	(-4.914, 2.177)
Unequal	-.77	31.43	.449	1.787	(-5.014, 2.276)

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 6				
PSYCHIATRIST	17	4.0588	5.178	1.256
RESEARCH	25	7.4400	8.124	1.625

Mean Difference = -3.3812

Levene's Test for Equality of Variances: F= 4.098 P= .050

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	-1.52	40	.137	2.230	(-7.890, 1.127)
Unequal	-1.65	39.89	.108	2.054	(-7.533, .770)

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 7				
PSYCHIATRIST	20	5.3500	8.368	1.871
RESEARCH	21	6.9048	7.880	1.720

Mean Difference = -1.5548

Levene's Test for Equality of Variances: F= .566 P= .456

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	-.61	39	.544	2.537	(-6.688, 3.579)
Unequal	-.61	38.53	.544	2.541	(-6.696, 3.587)

Variable	Number of Cases	Mean	SD	SE of Mean
OLD SITUATION CONDITION 8				
PSYCHIATRIST	17	11.8824	10.258	2.488
RESEARCH	20	10.6000	9.659	2.160

Mean Difference = 1.2824

Levene's Test for Equality of Variances: F= .107 P= .745

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	.39	35	.698	3.278	(-5.375, 7.939)
Unequal	.39	33.28	.700	3.295	(-5.422, 7.987)

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 1				
PSYCHIATRIST	21	3.2857	3.621	.790
RESEARCH	17	3.0000	2.979	.723

Mean Difference = .2857

Levene's Test for Equality of Variances: F= .705 P= .407

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	.26	36	.795	1.093	(-1.932, 2.504)
Unequal	.27	35.98	.791	1.071	(-1.886, 2.458)

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 2				
PSYCHIATRIST	17	2.8824	2.826	.685
RESEARCH	23	1.6522	1.921	.401

Mean Difference = 1.2302

Levene's Test for Equality of Variances: F= 1.221 P= .276

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	1.64	38	.109	.750	(-.289, 2.749)
Unequal	1.55	26.55	.133	.794	(-.399, 2.859)

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 3				
PSYCHIATRIST	16	3.5000	3.327	.832
RESEARCH	18	1.2222	.943	.222

Mean Difference = 2.2778

Levene's Test for Equality of Variances: F= 17.850 P= .000

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	2.79	32	.009	.817	(.612, 3.943)
Unequal	2.65	17.14	.017	.861	(.461, 4.094)

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 4				
PSYCHIATRIST	19	5.0000	4.320	.991
RESEARCH	18	4.8889	3.179	.749

Mean Difference = .1111

Levene's Test for Equality of Variances: F= 1.818 P= .186

t-test for Equality of Means					95%
Variances	t-value	df	2-Tail Sig	SE of Diff	CI for Diff
Equal	.09	35	.930	1.253	(-2.433, 2.655)
Unequal	.09	33.03	.929	1.243	(-2.417, 2.640)

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 5				
PSYCHIATRIST	19	7.2632	5.839	1.340
RESEARCH	17	5.7647	3.437	.834

Mean Difference = 1.4985

Levene's Test for Equality of Variances: F= 4.214 P= .048

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	.92	34	.362	1.622	(-1.799, 4.796)
Unequal	.95	29.64	.350	1.578	(-1.725, 4.722)

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 6				
PSYCHIATRIST	20	8.5500	5.907	1.321
RESEARCH	18	5.9444	4.582	1.080

Mean Difference = 2.6056

Levene's Test for Equality of Variances: F= 2.117 P= .154

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	1.51	36	.141	1.729	(-.902, 6.114)
Unequal	1.53	35.28	.136	1.706	(-.859, 6.070)

t-tests for independent samples of GROUP

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 7				
PSYCHIATRIST	19	9.8947	9.146	2.098
RESEARCH	21	11.9048	7.348	1.603

Mean Difference = -2.0100

Levene's Test for Equality of Variances: F= .473 P= .496

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	-.77	38	.446	2.612	(-7.298, 3.278)
Unequal	-.76	34.55	.452	2.641	(-7.372, 3.352)

Variable	Number of Cases	Mean	SD	SE of Mean
NEW SITUATION CONDITION 8				
PSYCHIATRIST	19	12.7895	11.674	2.678
RESEARCH	17	10.5882	5.269	1.278

Mean Difference = 2.2012

Levene's Test for Equality of Variances: F= 11.428 P= .002

t-test for Equality of Means Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	.71	34	.480	3.082	(-4.063, 8.466)
Unequal	.74	25.63	.465	2.967	(-3.900, 8.302)

Appendix 1.9: Pearson's Correlation between mean absolute errors and Z, the derivation of parameter A in the logistic map formula.

- - Correlation Coefficients - -

	Z	OLD	NEW	OLD#PSY	OLD#RES	NEW#PSY	NEW#RES
Z	1.0000	.8697	.9878	.7549	.8500	.9896	.9465
	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .	P= .005	P= .000	P= .030	P= .008	P= .000	P= .000
OLD	.8697	1.0000	.8394	.9318	.9543	.9126	.7427
	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .005	P= .	P= .009	P= .001	P= .000	P= .002	P= .035
NEW	.9878	.8394	1.0000	.7545	.8014	.9783	.9813
	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .009	P= .	P= .031	P= .017	P= .000	P= .000
OLD#PSY	.7549	.9318	.7545	1.0000	.7852	.8135	.6775
	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .030	P= .001	P= .031	P= .	P= .021	P= .014	P= .065
OLD#RES	.8500	.9543	.8014	.7852	1.0000	.8788	.7000
	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .008	P= .000	P= .017	P= .021	P= .	P= .004	P= .053
NEW#PSY	.9896	.9126	.9783	.8135	.8788	1.0000	.9204
	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .002	P= .000	P= .014	P= .004	P= .	P= .001
NEW#RES	.9465	.7427	.9813	.6775	.7000	.9204	1.0000
	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .035	P= .000	P= .065	P= .053	P= .001	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Appendix 1.10: Signed errors of the participants in the judgement trials

		Condition								
	Subject No.:	1	2	3	4	5	6	7	8	Means
Psychiatrist Group	1	0	-16	-1	3	4	0	-4	-18	-4.00
	2	0	-3	-5	-7	-8	-12	-4	-13	-6.50
	3	-5	-4	-7	-8	-8	-10	-19	-17	-9.75
	4	0	-1	-7	-10	-13	-13	-22	-24	-11.25
	5	-4	-2	-3	-6	-9	-7	-20	-17	-8.50
	6	-7	-5	-7	-10	-13	-16	-20	-25	-12.88
	7	-4	-11	-3	-4	-3	-8	-12	-25	-8.75
	8	-5	-3	-6	-4	-6	-10	-14	-40	-11.00
	9	-3	-8	-10	-2	-9	-3	-19	-53	-13.38
	10	-5	-4	-6	-7	-4	-8	-16	-23	-9.13
	11	-4	-4	-6	-7	-4	-13	-8	-34	-10.00
	12	-3	-3	-5	-4	-1	-2	-20	-8	-5.75
	13	7	7	5	4	4	-2	-48	-2	-3.13
	14	0	4	-3	0	-1	-10	-7	-13	-3.75
	15	-5	-3	-6	-7	-11	-10	-14	-18	-9.25
	16	-5	-1	-5	-4	-11	-8	-14	-18	-8.25
	17	-5	-4	-3	-7	-1	-5	-12	-10	-5.88
	18	-9	-5	-6	-13	-18	-15	-26	-43	-16.88
	19	-6	-5	-7	-13	-12	-15	-26	-33	-14.63
	20	-8	-5	-7	-14	-17	-18	-29	-26	-15.50
	21	-7	-4	-11	-9	-14	-16	-27	-23	-13.88
	22	-5	-4	-6	-7	-10	-14	-14	-14	-9.25
	23	-7	-5	-7	-10	-18	-25	-34	-48	-19.25
	24	-9	-7	-7	-9	-19	-17	-27	-31	-15.75
	25	-6	-4	-7	-11	-6	-10	-18	-18	-10.00
	26	0	0	-3	1	0	-2	-9	-38	-6.38
	27	-5	-1	-5	-7	4	-5	-17	2	-4.25
	28	-2	-1	-3	-7	-3	-6	-16	-13	-6.38
	29	0	-1	-3	-4	-6	-10	-14	-18	-7.00
	30	5	-1	-3	-7	4	0	-9	-3	-1.75
Means		-3.57	-3.47	-5.10	-6.33	-6.97	-9.67	-17.97	-22.13	-9.40
Research Group	31	-3	0	7	3	9	5	-9	-13	-0.13
	32	-5	-4	-5	-7	-6	-12	-14	-18	-8.88
	33	-5	-4	-6	-10	-9	-14	-20	-34	-12.75
	34	-8	-2	12	-1	-41	-15	-14	-33	-12.75
	35	-12	-8	-6	-6	-22	-28	-20	-40	-17.75
	36	-12	-2	0	-15	-6	-26	-28	-27	-14.50
	37	-7	-2	-1	-5	-11	-10	-16	-17	-8.63
	38	-5	1	-3	-5	-6	-10	-12	-21	-7.63
	39	4	2	-6	-21	3	0	1	-25	-5.25
	40	-6	-2	-6	-8	-6	-5	-9	-10	-6.50
	41	-7	-3	-4	0	-6	-5	-4	-38	-8.38
	42	-7	-1	-6	-7	-6	-10	-14	-21	-9.00
	43	-5	-1	-3	3	2	0	1	-36	-4.88
	44	8	1	-6	-17	-4	-5	-16	-12	-6.38
	45	-4	-1	-8	-7	-8	-10	-14	-18	-8.75
	46	-5	-4	-5	-7	-10	-10	-16	-13	-8.75
	47	-2	-1	-6	-2	2	0	-4	-18	-3.88
	48	-6	-4	-7	-7	-4	-13	-15	-16	-9.00
	49	-7	-4	-5	-11	-11	-16	-20	-33	-13.38
	50	-5	-1	-13	-7	-21	-30	-29	-23	-16.13
	51	-5	-1	-6	-4	-8	-12	-14	-43	-11.63
	52	-8	-5	-6	-7	-10	-13	-18	-21	-11.00
	53	-11	-5	-6	-6	-22	-12	-20	-31	-14.13
	54	-9	-3	-9	-17	-21	-22	-26	-31	-17.25
	55	-6	-7	-4	-5	-12	-15	-21	-38	-13.50
	56	-1	-1	-4	-27	-25	-30	-42	-48	-22.25
	57	0	-4	-5	-4	-6	-30	-9	-23	-10.13
	58	-6	-4	-6	-7	-6	-15	-16	-8	-8.50
	59	-5	-1	-3	-5	-11	-13	-19	-21	-9.75
	60	0	-1	-13	-7	4	-5	-4	-18	-5.50
Means		-5.00	-2.40	-4.63	-7.53	-9.27	-12.70	-15.37	-24.93	-10.23

Appendix 1.11: Absolute errors of the participants in the judgement trials

		Condition								
	Subject No.:	1	2	3	4	5	6	7	8	Means
Psychiatrist Group	1	0	16	1	3	4	0	4	18	5.75
	2	0	3	5	7	8	12	4	13	6.50
	3	5	4	7	8	8	10	19	17	9.75
	4	0	1	7	10	13	13	22	24	11.25
	5	4	2	3	6	9	7	20	17	8.50
	6	7	5	7	10	13	16	20	25	12.88
	7	4	11	3	4	3	8	12	25	8.75
	8	5	3	6	4	6	10	14	40	11.00
	9	3	8	10	2	9	3	19	53	13.38
	10	5	4	6	7	4	8	16	23	9.13
	11	4	4	6	7	4	13	8	34	10.00
	12	3	3	5	4	1	2	20	8	5.75
	13	7	7	5	4	4	2	48	2	9.88
	14	0	4	3	0	1	10	7	13	4.75
	15	5	3	6	7	11	10	14	18	9.25
	16	5	1	5	4	11	8	14	18	8.25
	17	5	4	3	7	1	5	12	10	5.88
	18	9	5	6	13	18	15	26	43	16.88
	19	6	5	7	13	12	15	26	33	14.63
	20	8	5	7	14	17	18	29	26	15.50
	21	7	4	11	9	14	16	27	23	13.88
	22	5	4	6	7	10	14	14	14	9.25
	23	7	5	7	10	18	25	34	48	19.25
	24	9	7	7	9	19	17	27	31	15.75
	25	6	4	7	11	6	10	18	18	10.00
	26	0	0	3	1	0	2	9	38	6.63
	27	5	1	5	7	4	5	17	2	5.75
	28	2	1	3	7	3	6	16	13	6.38
	29	0	1	3	4	6	10	14	18	7.00
	30	5	1	3	7	4	0	9	3	4.00
Means		4.37	4.20	5.43	6.87	8.03	9.67	17.97	22.27	9.85
Research Group	31	3	0	7	3	9	5	9	13	6.13
	32	5	4	5	7	6	12	14	18	8.88
	33	5	4	6	10	9	14	20	34	12.75
	34	8	2	12	1	41	15	14	33	15.75
	35	12	8	6	6	22	28	20	40	17.75
	36	12	2	0	15	6	26	28	27	14.50
	37	7	2	1	5	11	10	16	17	8.63
	38	5	1	3	5	6	10	12	21	7.88
	39	4	2	6	21	3	0	1	25	7.75
	40	6	2	6	8	6	5	9	10	6.50
	41	7	3	4	0	6	5	4	38	8.38
	42	7	1	6	7	6	10	14	21	9.00
	43	5	1	3	3	2	0	1	36	6.38
	44	8	1	6	17	4	5	16	12	8.63
	45	4	1	8	7	8	10	14	18	8.75
	46	5	4	5	7	10	10	16	13	8.75
	47	2	1	6	2	2	0	4	18	4.38
	48	6	4	7	7	4	13	15	16	9.00
	49	7	4	5	11	11	16	20	33	13.38
	50	5	1	13	7	21	30	29	23	16.13
	51	5	1	6	4	8	12	14	43	11.63
	52	8	5	6	7	10	13	18	21	11.00
	53	11	5	6	6	22	12	20	31	14.13
	54	9	3	9	17	21	22	26	31	17.25
	55	6	7	4	5	12	15	21	38	13.50
	56	1	1	4	27	25	30	42	48	22.25
	57	0	4	5	4	6	30	9	23	10.13
	58	6	4	6	7	6	15	16	8	8.50
	59	5	1	3	5	11	13	19	21	9.75
	60	0	1	13	7	4	5	4	18	6.50
Means		5.80	2.67	5.90	7.93	10.60	13.03	15.50	24.93	10.80

Appendix 1.12: A mixed ANOVA of the judgement trials for the overall group and condition effects

```
* * * * * * Analysis of Variance * * * * * *

        60 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         2 non-empty cells.

         1 design will be processed.
```

```
* * * * * Analysis of Variance -- design  1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL         7510.97      58    129.50
GROUP                    107.35       1    107.35        .83       .366
```

```
* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'COND' Within-Subject Effect.

AVERAGED Tests of Significance for COND using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL         14431.66     406     35.55
COND                    19475.13       7   2782.16      78.27       .000
GROUP BY COND             445.83       7     63.69       1.79       .087
```

Appendix 1.13: One-way ANOVA for each condition in the judgement trials

- - - - - O N E W A Y - - - - -

Variable COND1
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	30.8167	30.8167	3.7568	.0575
Within Groups	58	475.7667	8.2029		
Total	59	506.5833			

- - - - - O N E W A Y - - - - -

Variable COND2
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	35.2667	35.2667	4.8764	.0312
Within Groups	58	419.4667	7.2322		
Total	59	454.7333			

- - - - - O N E W A Y - - - - -

Variable COND3
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	3.2667	3.2667	.4808	.4908
Within Groups	58	394.0667	6.7943		
Total	59	397.3333			

- - - - - O N E W A Y - - - - -

Variable COND4
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	17.0667	17.0667	.7115	.4024
Within Groups	58	1391.3333	23.9885		
Total	59	1408.4000			

- - - - - O N E W A Y - - - - -

Variable COND5
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	98.8167	98.8167	1.8902	.1745
Within Groups	58	3032.1667	52.2787		
Total	59	3130.9833			

- - - - - O N E W A Y - - - - -

Variable COND6
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	170.0167	170.0167	3.0233	.0874
Within Groups	58	3261.6333	56.2351		
Total	59	3431.6500			

- - - - - O N E W A Y - - - - -

Variable COND7
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	91.2667	91.2667	1.1068	.2971
Within Groups	58	4782.4667	82.4563		
Total	59	4873.7333			

- - - - - O N E W A Y - - - - -

Variable COND8
By Variable GROUP

Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	106.6667	106.6667	.7558	.3882
Within Groups	58	8185.7333	141.1333		
Total	59	8292.4000			

Appendix 1.14: Mean absolute errors of the 30 psychiatrist group participants in the treatment trials

Condition 1		Condition 2		Condition 3		Condition 4		Condition 5		Condition 6		Condition 7		Condition 8	
First	Last	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last
3.75	7.50	10.75	23.50	10.75	18.50	4.25	9.25	4.25	1.00	5.00	15.00	10.50	9.00	13.00	13.00
2.00	1.00	1.00	1.00	1.00	1.00	13.50	10.75	10.50	14.25	5.75	2.00	6.00	13.50	10.75	5.50
5.75	1.25	3.00	1.00	4.75	1.25	9.50	10.00	8.50	9.50	14.00	16.00	14.50	9.75	20.25	26.75
8.25	3.00	1.00	1.00	16.50	10.50	13.75	12.25	11.25	14.25	18.50	20.00	24.75	25.00	31.25	35.25
2.50	1.25	1.75	1.00	3.50	2.25	6.00	5.00	5.25	4.00	11.75	10.75	10.00	8.00	20.00	27.00
4.75	1.25	3.50	3.00	6.50	4.75	8.00	9.25	11.00	7.50	17.50	18.25	19.50	19.00	19.50	14.00
5.00	2.25	4.50	3.00	3.00	3.00	3.00	1.00	7.25	4.75	14.75	15.00	10.75	9.25	16.50	16.00
3.50	2.50	1.00	1.00	4.50	3.00	3.00	2.00	7.50	6.75	9.50	8.00	15.50	11.50	21.00	17.50
0.00	0.00	3.50	2.25	10.50	10.75	3.00	1.00	8.50	9.00	8.75	6.50	9.25	1.00	1.50	1.00
5.00	0.00	2.00	2.00	1.50	0.00	2.00	2.00	4.50	4.00	9.00	6.25	16.25	14.00	19.50	12.25
1.75	1.00	0.50	0.00	9.50	4.75	3.25	2.00	6.75	3.25	4.50	3.75	9.00	5.00	9.25	13.50
6.75	3.75	3.75	2.50	7.00	7.50	8.00	8.50	18.75	13.50	11.25	9.50	19.00	16.25	21.00	25.25
3.25	3.00	12.50	10.00	9.50	2.00	3.00	3.00	3.75	3.00	5.00	12.00	3.75	1.00	5.75	2.00
0.00	5.00	13.00	11.25	2.00	2.00	2.00	2.00	5.50	7.00	8.75	10.50	11.50	1.75	18.25	17.50
5.75	5.50	5.50	3.75	7.50	6.00	14.50	8.75	14.75	16.75	19.75	20.00	16.50	18.50	19.25	29.25
2.50	3.50	13.75	12.50	3.00	3.00	10.00	15.25	11.00	6.50	12.75	16.00	17.75	10.50	20.75	22.25
5.75	0.25	1.00	1.00	1.50	0.00	3.50	1.00	1.00	1.00	5.50	2.50	13.50	6.25	14.25	5.25
11.00	10.00	4.75	4.00	6.50	7.00	11.50	13.75	13.50	14.25	19.00	20.00	15.25	44.50	33.00	27.00
7.00	3.25	4.75	3.25	6.25	1.75	11.00	9.50	13.25	12.75	19.50	20.75	23.00	25.00	26.50	27.00
8.50	5.50	4.25	5.00	9.00	7.75	11.75	11.75	15.75	15.50	18.25	19.50	23.25	23.75	29.50	28.00
10.00	2.50	5.25	3.50	9.25	8.25	9.50	12.75	12.00	13.75	17.50	21.25	24.50	25.25	28.75	31.25
9.25	8.25	3.00	3.00	6.50	5.75	9.75	9.25	12.25	12.00	16.25	15.25	17.50	14.00	31.00	26.25
9.00	9.25	5.00	4.50	17.50	32.75	12.25	12.25	30.25	20.25	28.75	13.25	23.75	22.50	36.00	32.75
6.25	3.75	4.00	4.00	5.00	8.50	11.00	10.50	15.00	16.25	19.25	19.75	17.00	25.00	33.00	39.00
6.50	4.00	3.75	2.75	3.00	2.25	2.00	2.00	6.00	1.00	17.50	15.75	9.75	13.75	25.00	25.50
2.00	1.00	3.00	2.00	13.00	13.00	3.25	3.50	1.75	2.00	2.50	1.25	2.25	1.00	43.75	19.00
3.50	3.00	2.25	1.00	2.50	2.00	5.75	2.00	4.75	16.00	6.25	4.50	4.00	1.75	8.00	1.50
3.00	1.00	0.75	0.00	5.25	3.00	7.50	5.75	5.75	5.00	5.00	6.75	6.25	3.75	26.25	23.50
5.50	5.00	1.00	1.00	3.00	3.25	5.50	2.00	10.50	6.00	13.75	14.25	19.00	15.25	25.50	22.00
0.00	0.00	1.75	1.00	2.00	2.00	8.00	2.00	3.00	1.00	6.25	3.75	7.75	3.25	10.50	2.50
4.93	3.28	4.18	3.83	6.38	5.92	7.30	6.67	9.46	8.73	12.39	12.27	14.04	13.27	21.28	19.62

Appendix 1.15: A simple ANOVA with repeated measures of the treatment trials

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * Analysis of Variance * * * * * *

        30 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         1 non-empty cell.

         1 design will be processed.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * Analysis of Variance -- design  1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL        8228.97      29    283.76
CONSTANT              44271.37       1  44271.37     156.02       .000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'COND' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL        8538.71     203     42.06
COND                  13354.29       7   1907.76      45.36       .000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'ORDER' Within-Subject Effect.

Tests of Significance for T9 using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL         337.28      29     11.63
ORDER                    76.56       1     76.56       6.58       .016

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'COND BY ORDER' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL        1940.07     203      9.56
COND BY ORDER .          33.94       7      4.85        .51       .828
```

Appendix 1.16: Pearson's Correlation between mean absolute errors in the both prediction trials and judgement trials for both experimental group.

- - Correlation Coefficients - -

	JUDG@PSY	JUDG@RES	PRED@PSY	PRED@RES
JUDG@PSY	1.0000	.3803	.5827	.4495
	(30)	(30)	(30)	(30)
	P= .	P= .038	P= .001	P= .013
JUDG@RES	.3803	1.0000	.4285	.1631
	(30)	(30)	(30)	(30)
	P= .038	P= .	P= .018	P= .389
PRED@PSY	.5827	.4285	1.0000	.4175
	(30)	(30)	(30)	(30)
	P= .001	P= .018	P= .	P= .022
PRED@RES	.4495	.1631	.4175	1.0000
	(30)	(30)	(30)	(30)
	P= .013	P= .389	P= .022	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Appendix Two

Raw data and Additional statistics for Experiment Two

APPENDIX TWO

Table of Contents:

Appendix 2.15: One sample Chi-squared on the performance in questionnaire for each question

Appendix 2.1: Absolute errors of the participants in the prediction trials

Psychiatrist Group

	Old Situation								New Situation							
	Condition								Condition							
Subject No.:	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	0	0	1	1	6	8	0	6	4	1	7	1	12	1	2	32
2	1	0	4	2	1	5	10	4	3	4	8	5	2	6	1	2
3	7	12	4	7	1	1	2	25	2	3	1	3	1	8	7	27
4	0	1	1	3	1	2	7	13	1	0	1	7	2	0	12	36
5	0	2	1	0	11	5	7	1	0	2	5	6	0	5	19	6
6	2	1	1	20	3	14	1	7	1	0	1	11	10	3	3	8
7	18	5	2	1	0	7	4	17	0	2	2	12	8	16	12	22
8	1	1	2	0	2	2	11	9	0	2	0	1	7	0	6	11
9	2	1	9	5	10	5	4	31	2	0	3	7	3	27	31	14
10	1	1	3	1	10	2	3	13	2	0	1	6	4	1	14	10
11	12	1	1	19	2	0	1	11	1	1	2	5	10	13	7	18
12	0	0	3	3	7	6	3	1	3	1	3	0	13	1	4	7
13	1	1	1	3	1	11	2	4	2	1	2	5	6	9	5	3
14	11	0	0	1	0	6	10	17	3	4	1	0	6	1	6	29
15	0	1	1	1	0	2	0	1	1	1	5	4	0	3	9	13
16	0	1	1	0	1	6	6	8	2	0	1	11	4	1	6	3
17	3	5	17	8	12	4	37	0	2	0	6	5	30	10	14	12
18	4	1	13	0	2	9	6	9	3	1	13	1	21	6	8	2
19	0	0	2	2	14	1	24	4	1	2	2	4	9	6	1	10
20	1	2	1	0	1	0	3	15	4	0	0	12	9	4	19	5
21	0	1	0	0	3	6	0	3	0	0	4	6	12	18	4	14
22	0	1	1	6	1	4	7	8	6	1	1	4	7	2	8	5
23	0	0	1	2	1	0	3	8	2	0	2	3	18	5	9	26
24	1	0	3	0	0	4	13	6	1	0	0	4	5	6	12	0
	2.71	1.58	3.04	3.54	3.75	4.58	6.83	9.21	1.92	1.08	2.96	5.13	8.29	6.33	9.13	13.13

Research Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
25	1	0	1	4	3	1	6	0	1	1	2	1	2	2	3	10
26	8	3	0	2	27	3	9	30	0	8	0	3	8	8	1	20
27	4	10	1	0	1	0	2	7	3	3	0	0	3	6	4	10
28	3	0	0	3	0	13	4	11	1	1	0	1	5	1	7	28
29	0	1	1	0	8	20	18	1	0	4	1	2	3	9	13	3
30	1	0	1	33	3	2	4	10	3	2	7	4	5	13	14	33
31	5	5	1	1	1	3	3	28	0	2	7	3	3	14	27	11
32	1	1	1	5	6	1	11	9	1	1	0	0	6	2	17	4
33	1	1	15	20	5	7	3	20	2	1	1	10	2	9	12	19
34	0	0	1	4	0	7	9	1	1	1	3	6	14	4	10	0
35	8	3	0	4	2	2	0	24	1	1	2	8	14	8	3	4
36	0	0	1	1	3	6	16	4	1	1	2	5	10	6	7	6
37	1	0	0	1	0	12	13	2	1	1	1	1	1	11	9	6
38	4	3	0	3	0	21	18	19	8	4	1	1	18	19	18	28
39	0	0	1	0	0	5	0	1	5	1	3	2	2	2	4	2
40	0	1	2	1	0	5	2	4	2	0	1	1	12	2	1	15
41	0	2	7	20	10	1	19	30	1	1	9	6	10	10	4	28
42	1	1	2	4	1	13	4	24	0	2	2	2	7	15	1	3
43	1	0	12	12	11	14	17	19	1	1	3	9	3	15	13	11
44	1	1	2	0	0	2	0	17	2	2	2	5	5	2	15	35
45	0	0	0	3	2	1	1	0	0	1	3	6	11	10	4	15
46	0	0	1	7	1	2	5	10	2	1	2	4	14	0	1	17
47	0	1	1	2	0	0	3	5	3	1	3	6	19	2	1	9
48	1	0	1	0	2	4	15	10	1	0	3	2	7	2	18	7
	1.71	1.38	2.17	5.42	3.58	6.04	7.58	11.92	1.67	1.71	2.42	3.67	7.67	7.17	8.63	13.50

Trainee Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
49	0	0	1	5	3	5	1	1	7	2	2	5	19	9	0	22
50	2	1	4	3	5	8	2	4	0	1	2	4	4	5	23	0
51	5	18	1	1	0	1	2	9	1	2	2	1	4	7	1	12
52	2	0	6	2	3	10	17	7	2	0	4	7	3	19	7	34
53	0	2	1	0	11	22	15	1	1	2	2	0	3	0	1	3
54	1	2	2	21	3	4	1	3	2	3	2	9	14	20	3	9
55	1	3	1	2	0	8	1	2	2	2	9	3	11	5	7	1
56	0	0	2	1	11	5	11	1	2	3	1	5	2	3	15	16
57	0	4	6	18	8	16	10	26	1	0	6	11	8	23	3	3
58	1	0	3	3	3	2	3	13	1	0	4	5	24	1	9	14
59	8	2	1	13	1	0	9	22	4	4	2	17	12	11	11	2
60	0	0	1	5	0	4	1	4	5	1	0	6	5	21	21	0
61	1	0	0	1	0	5	0	7	0	1	1	9	12	1	1	13
62	3	2	0	1	2	5	10	21	5	1	1	5	0	2	13	23
63	1	1	0	2	2	3	1	1	2	3	3	1	6	0	7	7
64	2	0	1	1	1	10	9	4	1	1	3	10	7	2	8	20
65	2	2	0	16	12	6	27	22	1	5	1	9	20	1	1	18
66	7	0	2	9	1	17	10	19	2	1	1	1	4	21	15	3
67	0	0	5	13	7	5	23	9	1	1	2	5	5	6	2	10
68	4	1	2	0	1	5	5	20	11	1	9	7	9	11	4	30
69	0	2	0	0	1	2	0	1	1	0	0	9	1	2	6	29
70	0	0	1	0	4	3	3	11	1	0	9	2	5	1	14	4
71	0	2	2	1	8	0	2	6	5	3	3	2	1	4	2	6
72	1	1	0	0	2	4	15	5	1	0	1	1	7	1	2	4
	1.71	1.79	1.75	4.92	3.71	6.25	7.42	9.13	2.46	1.54	2.92	5.58	7.75	7.33	7.33	11.79

Appendix 2.2: A simple ANOVA of the absolute errors of each participant in each trial in Appendix 2.1 with repeated measures on situation and condition

```
* * * * * * Analysis of Variance * * * * * *

        72 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         3 non-empty cells.

         1 design will be processed.
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL         4879.97      69     70.72
GROUP                      8.53       2      4.26        .06       .942
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'SIT' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL         2056.47      69     29.80
SIT                      431.45       1    431.45      14.48       .000
GROUP BY SIT              27.90       2     13.95        .47       .628
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares
Source of Variation          SS      DF        MS          F  Sig of F

WITHIN+RESIDUAL        16547.47     483     34.26
COND                   11275.89       7   1610.84      47.02       .000
GROUP BY COND            236.83      14     16.92        .49       .937
```

---

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'SIT BY COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	12514.05	483	25.91		
SIT BY COND	590.87	7	84.41	3.26	.002
GROUP BY SIT BY COND	143.77	14	10.27	.40	.976

Appendix 2.3: Paired sample t-test analyses of the absolute errors trial in Appendix 2.1 across both situations in each of the eight conditions

Condition 1

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	-.017	.885	2.0139	1.996	.235
OLD				2.0417	3.282	.387

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
-.0278	3.871	.456	-.06	71	.952
95% CI (-.938, .882)					

Condition 2

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	.229	.053	1.4444	1.452	.171
OLD				1.5833	2.852	.336

Paired Differences			t-value	df	2-tail Sig
Mean	SD	SE of Mean			
-.1389	2.889	.340	-.41	71	.685
95% CI (-.818, .540)					

Condition 3

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	.349	.003	2.7639	2.688	.317
OLD				2.3194	3.410	.402

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
.4444	3.528	.416	1.07	71	.289
95% CI (-.385, 1.274)					

Condition 4

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	.349	.003	4.7917	3.576	.421
OLD				4.6250	6.687	.788

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
.1667	6.389	.753	.22	71	.825
95% CI (-1.335, 1.668)					

Condition 5

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	.028	.817	7.9028	6.194	.730
OLD				3.6806	4.723	.557

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
4.2222	7.684	.906	4.66	71	.000
95% CI (2.416, 6.028)					

Condition 6

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	.290	.014	6.9444	6.569	.774
OLD				5.6250	5.163	.609

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
1.3194	7.083	.835	1.58	71	.118
95% CI (-.345, 2.984)					

Condition 7

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	.028	.815	8.3611	6.777	.799
OLD				7.2778	7.444	.877

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
1.0833	9.925	1.170	.93	71	.357
95% CI (-1.249, 3.416)					

Condition 8

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
NEW	72	.275	.019	12.8056	10.217	1.204
OLD				10.0833	8.679	1.023

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
2.7222	11.444	1.349	2.02	71	.047
95% CI (.032, 5.412)					

Appendix 2.4: Simple ANOVAs of the absolute errors of each participant in each trial in Appendix 2.1 with repeated measures on situation and condition for each experimental group

Psychiatrist Group

* * * * * * Analysis of Variance * * * * * *

24 cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
1 non-empty cell.

1 design will be processed.

* * * * * Analysis of Variance -- design 1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1420.75	23	61.77		
CONSTANT	10385.44	1	10385.44	168.13	.000

* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'SIT' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	424.93	23	18.48		
SIT	242.25	1	242.25	13.11	.001

* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	5620.94	161	34.91		
COND	3501.37	7	500.20	14.33	.000

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'SIT BY COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	4265.50	161	26.49		
SIT BY COND	329.81	7	47.12	1.78	.095

- -

Research Group

* * * * * * A n a l y s i s o f V a r i a n c e * * * * * *

24 cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
1 non-empty cell.

1 design will be processed.

- -

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	2117.50	23	92.07		
CONSTANT	11147.82	1	11147.82	121.09	.000

- -

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'SIT' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	649.48	23	28.24		
SIT	65.84	1	65.84	2.33	.140

- -

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	5946.19	161	36.93		
COND	4862.00	7	694.57	18.81	.000

- -

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'SIT BY COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	3822.79	161	23.74		
SIT BY COND	231.39	7	33.06	1.39	.212

- -

Trainee Group

* * * * * * A n a l y s i s o f V a r i a n c e * * * * * *

24 cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
1 non-empty cell.

1 design will be processed.

- -

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1341.73	23	58.34		
CONSTANT	10427.09	1	10427.09	178.74	.000

- -

* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'SIT' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	982.06	23	42.70		
SIT	151.25	1	151.25	3.54	.073

- -

* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	4980.34	161	30.93		
COND	3149.35	7	449.91	14.54	.000

- -

* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'SIT BY COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	4425.75	161	27.49		
SIT BY COND	173.43	7	24.78	.90	.507

- -

Appendix 2.5: Absolute errors of the participants in the judgement trials

Psychiatrist Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
1	3	6	3	5	8	11	14	8	4	2	0	8	7	6	4	8
2	7	4	11	2	11	15	4	18	6	4	13	9	11	25	29	18
3	6	10	7	6	6	2	11	14	1	2	5	6	11	16	3	26
4	10	3	5	4	12	10	12	32	3	2	3	8	16	8	16	33
5	0	1	1	3	0	10	6	2	2	3	2	4	1	4	1	11
6	4	5	8	7	15	16	16	29	13	4	7	10	8	10	23	18
7	4	8	2	1	1	8	4	18	8	2	5	8	11	9	4	8
8	5	1	6	11	15	12	23	27	5	10	3	7	15	16	19	17
9	11	4	3	7	11	15	29	20	5	4	7	7	14	10	29	23
10	4	2	5	7	8	11	13	14	4	1	3	5	4	10	14	27
11	5	1	2	7	6	4	9	3	5	3	5	7	9	14	32	21
12	3	1	5	8	4	9	12	19	3	2	3	7	10	3	14	16
13	3	1	3	7	1	15	9	18	20	4	3	11	2	5	20	22
14	0	1	7	2	26	5	14	16	2	3	4	7	4	0	7	11
15	20	2	2	2	1	4	3	10	3	3	3	8	4	1	8	5
16	1	1	1	1	4	14	3	10	6	0	2	4	13	5	25	11
17	5	1	6	17	21	19	26	2	2	1	7	9	14	12	19	18
18	13	4	7	11	15	19	22	29	7	4	6	11	14	16	16	27
19	0	1	11	6	13	19	6	16	0	1	4	7	8	4	29	53
20	1	2	4	2	5	10	11	19	1	2	2	10	7	11	11	15
21	0	3	1	3	6	15	1	3	2	3	3	7	13	0	19	2
22	2	4	2	7	6	10	14	13	2	1	5	7	15	5	28	24
23	3	2	3	1	5	4	17	11	0	2	4	6	6	8	2	20
24	9	2	7	12	15	21	25	23	6	9	3	7	5	4	25	18
	4.96	2.92	4.67	5.79	8.96	11.58	12.67	15.58	4.58	3.00	4.25	7.50	9.25	8.42	16.54	18.83

Research Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
25	2	2	3	19	7	24	30	15	4	3	2	7	8	30	29	3
26	3	1	13	15	21	24	10	18	7	1	5	10	28	5	12	36
27	10	14	23	27	1	10	14	8	0	0	5	4	1	5	44	43
28	11	5	6	6	14	20	16	23	4	3	5	8	6	11	17	34
29	0	1	2	3	4	10	34	2	2	4	2	4	15	12	14	14
30	3	6	6	11	15	20	16	29	6	2	23	12	6	5	19	10
31	5	9	3	5	6	4	14	13	9	3	5	2	1	3	9	15
32	5	1	7	10	11	15	23	27	3	1	7	11	8	12	18	26
33	5	1	14	9	6	19	29	43	2	4	7	18	11	10	32	20
34	6	2	7	7	11	13	14	18	4	3	2	5	10	8	12	38
35	5	1	10	2	26	18	34	38	13	3	9	8	12	19	29	18
36	3	7	4	2	3	7	6	18	3	2	2	8	9	4	31	10
37	2	3	5	4	1	10	16	15	3	3	15	7	11	8	17	20
38	0	1	4	2	14	12	9	31	2	3	4	7	19	13	9	3
39	1	7	3	2	1	13	2	10	4	4	3	10	9	10	8	7
40	1	1	1	1	3	17	6	12	6	3	4	3	9	5	14	18
41	0	9	3	2	6	17	9	48	2	3	3	7	10	15	16	2
42	7	4	6	11	15	17	22	30	7	1	5	11	13	11	14	27
43	0	1	4	7	4	19	54	18	0	1	3	32	14	50	10	26
44	15	11	2	3	4	10	9	16	3	4	3	2	6	0	9	3
45	2	2	3	3	8	17	6	8	1	2	2	8	13	10	17	8
46	1	2	2	2	16	5	14	28	3	3	4	3	16	5	19	26
47	3	3	1	0	4	4	12	10	2	3	2	7	3	20	2	14
48	7	2	6	9	20	17	14	18	5	2	5	6	26	0	34	48
	4.04	4.00	5.75	6.75	9.21	14.25	17.21	20.67	3.96	2.54	5.29	8.33	11.00	11.29	18.13	19.54

Trainee Group

Subject No.:	Old Situation Condition 1	2	3	4	5	6	7	8	New Situation Condition 1	2	3	4	5	6	7	8
49	3	3	3	4	7	12	13	19	5	3	1	8	7	8	7	3
50	5	3	5	7	12	15	27	23	6	3	5	8	11	5	9	13
51	27	6	20	5	7	3	12	15	0	11	4	5	10	10	4	26
52	0	3	5	3	13	18	15	22	4	2	6	8	6	20	4	3
53	0	1	1	3	2	4	30	2	2	4	4	4	9	15	14	18
54	7	3	6	4	12	18	20	29	5	5	7	7	7	24	19	30
55	5	7	2	1	1	9	4	18	6	2	5	8	10	10	4	8
56	6	11	3	7	6	15	19	23	0	16	3	6	11	30	19	23
57	7	4	6	7	11	17	27	23	3	4	6	7	14	15	29	13
58	2	14	0	7	8	25	16	20	5	6	8	3	8	28	20	36
59	4	3	2	2	14	4	9	38	8	4	6	10	11	18	18	30
60	3	1	5	8	4	7	10	18	3	2	2	18	10	4	0	10
61	1	1	3	3	1	13	14	14	4	3	2	2	7	6	13	23
62	7	1	2	7	26	10	14	18	7	4	4	7	6	15	14	23
63	1	2	2	1	1	10	6	8	2	4	1	11	4	15	3	18
64	5	2	0	1	8	11	1	46	6	1	0	20	20	8	32	13
65	0	4	2	17	36	17	34	33	13	1	18	17	8	30	29	33
66	9	3	6	10	13	24	20	23	7	4	4	9	10	11	25	31
67	0	1	10	7	11	11	4	13	3	4	5	7	6	10	9	27
68	1	0	3	2	6	12	11	18	2	5	2	9	8	10	12	18
69	20	1	3	3	6	15	34	8	5	10	7	5	1	2	14	23
70	1	1	2	2	1	5	5	18	4	4	5	7	15	10	14	13
71	4	3	2	0	4	4	15	12	0	3	4	7	6	8	0	10
72	5	3	5	7	4	17	4	18	5	2	5	5	1	3	27	10
	5.13	3.38	4.08	4.92	8.92	12.33	15.17	19.96	4.38	4.46	4.75	8.25	8.58	13.13	14.13	18.96

Appendix 2.6: A simple ANOVA of the absolute errors of each participant in each trial in Appendix 2.5 with repeated measures on situation and condition

```
* * * * * * Analysis of Variance * * * * * *

        72 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         3 non-empty cells.

         1 design will be processed.
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL         10311.06      69    149.44
GROUP                     378.34       2    189.17      1.27      .288
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'SIT' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL          2020.81      69     29.29
SIT                        19.27       1     19.27       .66      .420
GROUP BY SIT               38.23       2     19.12       .65      .524
```

```
* * * * * Analysis of Variance -- design 1 * * * * *

Tests involving 'COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL         18998.46     483     39.33
COND                    32113.84       7   4587.69    116.63      .000
GROUP BY COND             535.64      14     38.26       .97      .480
```

---

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'SIT BY COND' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	19318.13	483	40.00		
SIT BY COND	349.99	7	50.00	1.25	.274
GROUP BY SIT BY COND	480.07	14	34.29	.86	.606

- -

Appendix 2.7: A simple ANOVA of the absolute errors of each participant with repeated measures on situation and mood state

```
* * * * * * Analysis of Variance * * * * * *

        72 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         3 non-empty cells.

         1 design will be processed.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * Analysis of Variance -- design  1 * * * * *

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL         2340.50      69     33.92
GROUP                     70.20       2     35.10      1.03      .361

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'SIT' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL          644.94      69      9.35
SIT                         .52       1       .52       .06      .814
GROUP BY SIT              12.30       2      6.15       .66      .521

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'MOOD' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares
Source of Variation          SS      DF        MS         F  Sig of F

WITHIN+RESIDUAL         2170.11     138     15.73
MOOD                    6438.88       2   3219.44    204.73      .000
GROUP BY MOOD             88.57       4     22.14      1.41      .235

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'SIT BY MOOD' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1247.66	138	9.04		
SIT BY MOOD	11.93	2	5.97	.66	.519
GROUP BY SIT BY MOOD	11.43	4	2.86	.32	.867

- -

Appendix 2.8: Pearson's Correlations of absolute errors between the three experimental groups

Old situation

- - Correlation Coefficients - -

	PSYCHIAT	RESEARCH	TRAINEE
PSYCHIAT	1.0000	.7152	.7722
	(24)	(24)	(24)
	P= .	P= .000	P= .000
RESEARCH	.7152	1.0000	.7336
	(24)	(24)	(24)
	P= .000	P= .	P= .000
TRAINEE	.7722	.7336	1.0000
	(24)	(24)	(24)
	P= .000	P= .000	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

New situation

- - Correlation Coefficients - -

	PSYCHIAT	RESEARCH	TRAINEE
PSYCHIAT	1.0000	.3137	.2773
	(24)	(24)	(24)
	P= .	P= .136	P= .190
RESEARCH	.3137	1.0000	.2801
	(24)	(24)	(24)
	P= .136	P= .	P= .185
TRAINEE	.2773	.2801	1.0000
	(24)	(24)	(24)
	P= .190	P= .185	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Overall

- - Correlation Coefficients - -

	PSYCHIAT	RESEARCH	TRAINEE
PSYCHIAT	1.0000	.5609	.5974
	(48)	(48)	(48)
	P= .	P= .000	P= .000
RESEARCH	.5609	1.0000	.5780
	(48)	(48)	(48)
	P= .000	P= .	P= .000
TRAINEE	.5974	.5780	1.0000
	(48)	(48)	(48)
	P= .000	P= .000	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Appendix 2.9: Pearson's Correlations of absolute errors between the three experimental groups

Old situation

- - Correlation Coefficients - -

	PSYCHIAT	RESEARCH	TRAINEE
PSYCHIAT	1.0000	.6019	.5713
	(24)	(24)	(24)
	P= .	P= .002	P= .004
RESEARCH	.6019	1.0000	.5710
	(24)	(24)	(24)
	P= .002	P= .	P= .004
TRAINEE	.5713	.5710	1.0000
	(24)	(24)	(24)
	P= .004	P= .004	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

New situation

- - Correlation Coefficients - -

	PSYCHIAT	RESEARCH	TRAINEE
PSYCHIAT	1.0000	.6490	.3756
	(24)	(24)	(24)
	P= .	P= .001	P= .070
RESEARCH	.6490	1.0000	.0587
	(24)	(24)	(24)
	P= .001	P= .	P= .785
TRAINEE	.3756	.0587	1.0000
	(24)	(24)	(24)
	P= .070	P= .785	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Overall

- - Correlation Coefficients - -

	PSYCHIAT	RESEARCH	TRAINEE
PSYCHIAT	1.0000	.6157	.4798
	(48)	(48)	(48)
	P= .	P= .000	P= .001
RESEARCH	.6157	1.0000	.3292
	(48)	(48)	(48)
	P= .000	P= .	P= .022
TRAINEE	.4798	.3292	1.0000
	(48)	(48)	(48)
	P= .001	P= .022	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Appendix 2.10: Pearson's Correlation between mean absolute errors and Z, the derivation of parameter A in the logistic map formula.

- - Correlation Coefficients - -

	Z	OLD	NEW	OLD# PSY	OLD# RES	OLD# TRA	NEW# PSY	NEW# RES	NEW# TRA
Z	1.0000	.9577	.9521	.9525	.9313	.9565	.9433	.9578	.9337
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .	P= .000	P= .000	P= .000	P= .001	P= .000	P= .000	P= .000	P= .001
OLD	.9577	1.0000	.9358	.9774	.9956	.9847	.9307	.9341	.9222
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .	P= .001	P= .000	P= .000	P= .000	P= .001	P= .001	P= .001
NEW	.9521	.9358	1.0000	.9310	.9247	.9162	.9951	.9934	.9908
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .001	P= .	P= .001	P= .001	P= .001	P= .000	P= .000	P= .000
OLD# PSY	.9525	.9774	.9310	1.0000	.9658	.9333	.9341	.9368	.8987
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .000	P= .001	P= .	P= .000	P= .001	P= .001	P= .001	P= .002
OLD# RES	.9313	.9956	.9247	.9658	1.0000	.9761	.9177	.9199	.9173
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .001	P= .000	P= .001	P= .000	P= .	P= .000	P= .001	P= .001	P= .001
OLD# TRA	.9565	.9847	.9162	.9333	.9761	1.0000	.9066	.9122	.9113
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .000	P= .001	P= .001	P= .000	P= .	P= .002	P= .002	P= .002
NEW# PSY	.9433	.9307	.9951	.9341	.9177	.9066	1.0000	.9833	.9808
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .001	P= .000	P= .001	P= .001	P= .002	P= .	P= .000	P= .000
NEW# RES	.9578	.9341	.9934	.9368	.9199	.9122	.9833	1.0000	.9743
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .000	P= .001	P= .000	P= .001	P= .001	P= .002	P= .000	P= .	P= .000
NEW# TRA	.9337	.9222	.9908	.8987	.9173	.9113	.9808	.9743	1.0000
	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)
	P= .001	P= .001	P= .000	P= .002	P= .001	P= .002	P= .000	P= .000	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Appendix 2.11: Pearson's Correlation between absolute errors in each trial and Z, the derivation of parameter A in the logistic map formula.

Old prediction trials

- - Correlation Coefficients - -

	ERROR	Z
ERROR	1.0000	.4203
	(576)	(576)
	P= .	P= .000
Z	.4203	1.0000
	(576)	(576)
	P= .000	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

New prediction trials

- - Correlation Coefficients - -

	ERROR	Z
ERROR	1.0000	.5144
	(576)	(576)
	P= .	P= .000
Z	.5144	1.0000
	(576)	(576)
	P= .000	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

All trials

- - Correlation Coefficients - -

	ERROR	Z
ERROR	1.0000	.4666
	(1152)	(1152)
	P= .	P= .000
Z	.4666	1.0000
	(1152)	(1152)
	P= .000	P= .

(Coefficient / (Cases) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Appendix 2.12: Mean absolute errors of the psychiatrist group participants in the treatment trials

Condition 1		Condition 2		Condition 3		Condition 4		Condition 5		Condition 6		Condition 7		Condition 8	
First	Last	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last
5.25	3.00	3.75	3.00	4.75	3.00	5.00	6.50	7.75	6.50	14.25	12.50	13.50	4.75	19.25	5.50
7.00	7.00	5.00	4.25	11.00	11.00	17.00	9.50	12.75	13.50	15.00	15.00	20.25	19.25	18.00	18.00
15.50	5.25	13.25	16.00	6.75	0.50	5.00	1.50	4.75	6.50	4.75	4.00	8.50	10.50	14.25	14.00
2.75	0.75	4.00	3.00	6.00	3.75	9.00	9.00	12.25	14.00	17.00	14.50	20.25	15.00	22.00	22.25
0.00	0.00	1.25	1.00	1.00	1.00	2.50	3.00	9.00	2.25	9.00	15.00	9.25	6.00	2.00	1.00
4.25	1.75	5.25	3.00	7.00	6.00	16.25	10.50	14.50	14.50	21.00	19.75	18.25	16.00	28.75	27.75
5.00	5.00	9.00	8.00	2.25	4.00	1.50	1.00	2.50	1.00	6.25	8.25	5.25	7.50	29.25	28.00
8.50	5.75	5.25	1.75	7.00	5.25	11.00	11.00	14.00	14.50	16.75	11.50	23.25	21.50	27.00	23.50
5.00	2.75	1.00	1.00	10.00	2.25	9.75	8.25	8.50	17.25	19.25	18.50	20.50	22.25	29.25	24.00
6.00	4.75	1.00	1.00	5.75	5.00	6.75	4.50	8.75	9.25	12.00	11.50	16.00	15.75	19.75	20.50
16.25	11.25	1.75	1.25	3.25	2.00	20.00	16.25	5.25	4.00	3.75	6.00	15.25	24.00	24.25	6.75
4.50	3.00	1.25	1.00	6.50	4.75	8.00	7.50	4.75	3.75	8.00	6.75	5.00	8.75	20.50	14.00
0.75	1.00	1.75	3.00	4.25	3.00	3.75	1.75	12.00	2.75	10.00	8.75	10.00	1.00	14.25	11.75
2.50	0.00	1.00	1.00	4.00	3.00	4.50	2.00	9.50	10.75	12.50	5.00	16.50	24.00	19.25	28.00
3.00	1.00	4.25	3.00	4.50	2.75	6.00	2.50	1.50	1.00	12.50	15.75	2.50	7.50	14.00	9.00
2.50	2.00	1.00	1.00	3.25	1.00	1.75	1.00	3.50	3.00	12.25	12.50	6.25	2.25	12.00	9.75
1.25	1.25	3.50	1.75	5.50	2.75	25.75	23.25	25.00	41.00	11.25	18.00	22.75	20.25	26.75	29.00
8.75	7.50	4.75	4.00	6.75	5.50	11.00	10.75	16.50	16.50	19.50	20.50	22.50	23.00	30.50	29.00
1.75	0.00	1.00	1.00	4.00	12.75	7.25	10.25	14.75	15.50	24.75	8.50	22.50	49.50	13.75	24.25
0.50	2.00	0.50	1.25	3.50	3.50	2.00	2.00	6.00	6.00	6.75	10.00	12.00	11.00	18.25	20.50
1.25	0.00	1.00	1.75	3.00	3.00	2.75	3.00	6.00	2.75	10.00	15.00	4.75	2.25	9.25	2.75
1.00	0.00	6.75	2.00	2.25	2.75	2.75	3.25	3.50	2.75	7.75	9.25	5.75	4.00	13.00	16.75
6.25	2.75	2.75	4.25	4.50	3.50	5.00	1.00	6.50	5.25	9.00	5.50	18.00	13.50	11.75	10.00
8.50	5.50	4.50	2.50	7.00	7.00	12.00	12.00	16.00	16.75	19.50	19.50	22.25	23.75	23.00	24.50

Appendix 2.13: A simple ANOVA with repeated measures of the treatment trials

```
* * * * * * Analysis of Variance * * * * * *

        24 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         1 non-empty cell.

         1 design will be processed.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


* * * * * Analysis of Variance -- design  1 * * * * *

Tests of Between-Subjects Effects.

 Tests of Significance for T1 using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

 WITHIN+RESIDUAL         5064.22      23    220.18
 CONSTANT               32860.63       1  32860.63    149.24      .000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'COND' Within-Subject Effect.

 AVERAGED Tests of Significance for VAR using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

 WITHIN+RESIDUAL         6788.90     161     42.17
 COND                    9991.28       7   1427.33     33.85      .000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


* * * * * Analysis of Variance -- design  1 * * * * *

Tests involving 'ORDER' Within-Subject Effect.

 Tests of Significance for T9 using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F

 WITHIN+RESIDUAL          236.29      23     10.27
 ORDER                     59.57       1     59.57      5.80      .024

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

* * * * * A n a l y s i s o f V a r i a n c e -- design 1 * * * * *

Tests involving 'COND BY ORDER' Within-Subject Effect.

AVERAGED Tests of Significance for VAR using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1504.15	161	9.34		
COND BY ORDER	63.27	7	9.04	.97	.457

- -

Appendix 2.14: Performance of participants in the questionnaire section

Psychiatrist				Researcher				Trainee			
Q.1	Q.2	Q.3	Q.4	Q.1	Q.2	Q.3	Q.4	Q.1	Q.2	Q.3	Q.4
1	1	0	1	0	1	1	0	1	0	0	0
1	1	0	0	0	1	0	0	1	1	0	0
1	0	1	0	1	1	1	1	0	1	0	0
1	0	0	1	0	0	0	0	1	1	1	0
1	1	1	0	1	0	1	0	1	1	1	1
1	0	0	0	1	0	0	0	1	1	0	0
1	1	0	0	1	1	0	0	1	1	0	0
1	0	1	1	1	1	1	0	1	1	0	0
1	0	0	0	1	1	1	0	1	0	1	0
1	0	0	0	1	1	0	1	1	1	0	0
1	1	0	0	1	0	1	0	0	0	0	0
1	0	0	0	1	0	0	0	1	1	1	0
1	1	0	0	1	1	0	0	1	1	1	0
1	0	0	0	1	1	1	0	1	0	0	0
1	0	0	0	1	1	0	0	1	0	0	0
1	0	0	0	1	1	0	0	1	1	0	0
1	0	1	0	1	0	0	0	1	0	0	0
0	1	1	0	1	0	1	0	1	1	0	0
1	0	0	0	1	1	0	0	1	1	1	0
1	1	0	0	1	1	0	1	1	0	0	0
1	1	1	0	1	1	0	0	1	0	0	0
1	0	0	0	1	0	0	0	1	1	0	0
1	0	0	0	1	1	0	1	1	1	1	0
1	1	0	0	1	1	1	0	1	0	0	1
23	10	6	3	21	16	9	4	22	15	7	2

Appendix 2.15: One sample Chi-squared on the performance in questionnaire for each question

- - - - - Chi-Square Test

QUEST1

Category	Cases Observed	Expected	Residual
1.00	23	22.00	1.00
2.00	21	22.00	-1.00
3.00	22	22.00	.00
Total	66		

Chi-Square	D.F.	Significance
.0909	2	.9556

- - - - - Chi-Square Test

QUEST2

Category	Cases Observed	Expected	Residual
1.00	10	13.67	-3.67
2.00	16	13.67	2.33
3.00	15	13.67	1.33
Total	41		

Chi-Square	D.F.	Significance
1.5122	2	.4695

- - - - - Chi-Square Test

QUEST3

Category	Cases Observed	Expected	Residual
1.00	6	7.33	-1.33
2.00	9	7.33	1.67
3.00	7	7.33	-.33
Total	22		

Chi-Square	D.F.	Significance
.6364	2	.7275

- - - - - Chi-Square Test

QUEST4

Category	Cases Observed	Expected	Residual
1.00	21	21.00	.00
2.00	20	21.00	-1.00
3.00	22	21.00	1.00
Total	63		

Chi-Square	D.F.	Significance
.0952	2	.9535

Bibliography

BIBLIOGRAPHY

Ayers, S. (1997). The application of chaos theory to Psychology. *Theory and Psychology*, Vol. 7(3): 373-398.

Berry, D. C. and Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36A, 209-231.

Buchner, A., Funke, J. and Berry, D. C. (1995). Negative correlations between control performance and verbalisable knowledge: Indicators for implicit learning in process control tasks? *The Quarterly Journal of Experimental Psychology*, 48A, 166-187.

Dienes, Z., and Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 848-862.

Evans, J. St. B. T. (1987). Beliefs and expectations as causes of judgmental bias. In George Wright & Peter Ayton (Eds), *Judgmental forecasting*, John Wiley & Sons Ltd.

Geddes, B. W. and Stevenson, R. J. (1997). The effects of learning goals on implicit and explicit learning, *The Quarterly Journal of Experimental Psychology*, 50A(4), 742-765.

Geisser, S. (1993). *Predictive inference: an introduction*, New York and London: Chapman & Hall

Harvey, N. (1990). Judgmental control of the behavior of a dynamical system. In K. J. Gilhooly, G. Erodos, M. Keane & R.Logie (Eds), *Lines of Thinking: Reflections on the Psychology of Thought*, pp. 337-352. New York: Wiley.

Harvey, N., Bolger, F. and McClelland, A. (1994). On the nature of expectations, *British Journal of Psychology*, 85, 203-229.

Harvey, N., Koehler, D.J. and Ayton, P. (1997). Judgments of decision effectiveness: Actor-observer differences in overconfidence, *Organizational behavior and human decision processes*, Vol.70, No.3, June, pp. 267-282.

Hayes, N. A., and Broadbent, D. E. (1988). Two modes of learning for interactive task. *Cognition*, 28, 249-276.

Koehler, D.J. and Harvey, N. (1997). Confidence judgements by actors and observers. *Journal of Behavioral Decision Making*, Vol. 10, pp. 221-242.

Lindsay, J. S. B. (1997). Psychiatry and chaos. *Medical hypotheses*, Vol.48, No.2, pp.149-159.

Mason, S. J. (1998). Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *International Journal of Climatology*, 1998, Vol.18, No.2, pp.147-164.

Montgomery, A. L. , Zarnowitz, V, Tsay, R. S. and Tiao, G. C. (1998). Forecasting the US unemployment rate. *Journal of the American Statistical Association*, 1998, Vol.93, No.442, pp.478-493.

Paulus, M. P., Geyer, M. A. and Braff, D. L. (1996). Use of methods from chaos theory to quantify a fundamental dysfunction in the behavioural organisation of schizophrenic patients, *American Journal of Psychiatry*, Vol.153, No. 5, pp. 714-717.

Phillips, L.D. (1987). On the adequacy of judgmental forecasting. In George Wright & Peter Ayton (Eds), *Judgmental forecasting*, John Wiley & Sons Ltd.

Reber, A. S. (1967). Implicit learning of artificial grammers. *Journal of verbal learning and verbal behaviour*, 5, 855-863.

Simon, H. A. and Lea, G. (1974). Problem solving and rule induction: a unified view. In L.W. Gregg (Ed.) *Knowledge and Cognition* (pp.105-128). Potomac, Maryland: Lawrence Erlbaum Associates.

Sontag, E.D. (1990). *Mathematical control theory: deterministic finite dimensional systems*, New York: Springer-Verlag.