The effects of learning goals on implicit and explicit learning

Geddes, Bruce Wynter

The Effects Of Learning Goals On Implicit And Explicit Learning

Bruce Wynter Geddes

Submitted for the qualification of

Doctor of Philosophy

University of Durham,

Psychology Department

June 1997

Abstract

This thesis, using Berry & Broadbent's (1984) computer-person interaction task, shows that three different learning goals result in three different learning modes. Experiment 1 demonstrated this effect: a pattern search goal resulted in explicit rule learning; a control task goal, as used in previous studies, resulted in instance learning where all instances are entered into a look-up table irrespective of whether the instance had been performed correctly or incorrectly; a dual goal, consisting of a combination of the last two goals, resulted in instance learning where only correct instances were entered into a look-up table. Experiment 2 refuted one explanation of the learning goal effect - it is not due to an indirect effect of altering the range of interactions that subjects see. Therefore, it must be due to a direct cognitive effect. Experiment 3 explored this direct effect showing that, in terms of Klahr and Dunbar's (1988) Dual Space model, a pattern search goal encourages the search primarily of rule space whereas a control task goal confines subjects to a search of instance space. The positive effect of self-explanations on both instance and rule learning was also demonstrated. Experiment 4 showed that subjects with the dual goal learn purely implicitly - all goal groups with a concurrent task of random number generation produced identical results to those of Experiment 1's dual goal group. Experiment 5 examined the learning goal effect on memory. Surprisingly, pattern search learners may still learn instances and dual goal subjects may still memorise instances on which they make errors. Control task learners' abilities are a simple reflection of their memories. Experiments 6a and 6b showed that only near transfer of learning occurs for control task instance learners. However, far transfer also occurs for pattern search learners, but only when the task transferred to is less complex, or of comparable complexity.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Rosemary Stevenson for all her kind help, support, intuition, patience and time. I would also like to thank the BBSRC for funding me through the last few years. A large note of gratitude goes out to Dr Oliver Davidson and Jan Lister for additional support, particularly during the beginning of my time as a postgraduate. I would also like to thank Andrew Bilecki for his balding support and the interest he gave, despite having to worry about higher concerns. Another note of gratitude goes to Tony Hall, without who's help I would probably never have embarked on the course I have taken (I am sorry I did not mention it before). A large lorry full of thanks goes to my Parents and sister for supporting me at all points in my education (I am not sure they realise how grateful I am). I have to also say a thank you to the crazy gang for fun and games and the rest of it. Last, but not least, a pinkly fab thank you, which is the only one appropriate for her and possibly the only one she would except, to Fi, who has come into my life like a train that can't find a station which is fine by me thank you.

CONTENTS

	PAGE No.
Table of Figures	iii
Table of Tables	vi
Declaration	vii
Opening Introduction	1
Chapter 1 - Experiment 1: Learning Goals and Implicit vs Explicit Learning	11
Introduction	12
Method	21
Results	25
Discussion	34
Chapter 2 - Experiment 2: Learning Goals - A Direct Influence On Cognitive Processes Or Just On Salience?	42
Introduction	43
Method	49
Results	53
Discussion	65
Chapter 3 - Experiment 3: Learning Goals and Explanations - How Do They Influence Cognitive Processes?	70
Introduction	71
Method	80
Results	82
Discussion	93
Chapter 4 - Experiment 4: Implicit Learning	103
Introduction	104
Method	110
Results	112
Discussion	120

Chapter 5 - Experiment 5: Learning Goals and Memory 128

Introduction 129

Method 133

Results 135

Discussion 141

Chapter 6 - Experiments 6a & 6b: The Rule/Instance Learning Distinction and Transfer of Learning. 152

Introduction 153

Experiment 6a : Method 162

Results 166

Discussion 176

Experiment 6b : Method 181

Results 182

Discussion 192

General Discussion 195

Closing Discussion 200

Appendix 1 - F tables and additional statistics for Experiment 1 228

Appendix 2 - F tables and additional statistics for Experiment 2 233

Appendix 3 - F tables and additional statistics for Experiment 3 236

Appendix 4 - F tables and additional statistics for Experiment 4 244

Appendix 5 - F tables and additional statistics for Experiment 5 248

Appendix 6 - F tables and additional statistics for Experiment 6a & 6b 251

Appendix 7 - Examples of answers to the rule description questions 261

Appendix 8 - Some transcribed concurrent verbalisations from Experiment 3 263

Appendix 9 - Contents of accompanying floppy disc. (It contains SPSS files of all the data) 265

Bibliography 266

Table of Figures

Figure number	Description	Page no.
1.1	Mean number of correct trials in the learning phase for the specific and dual goal groups. Data are shown for all 30 trials combined and for the first and second 15 trials.	26
1.2	Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.	27
1.3	Mean percentage of correct responses to each category of prediction questions for each group.	29
2.1	Mean number of correct trials in the test phase for the models. Data are shown for all 30 trials combined and for the first and second 15 trials.	55
2.2	Mean percentage of correct responses to each category of prediction question for the models.	56
2.3	State encounters for the two model groups.	59
2.4	Mean number of correct trials in the test phase for the sets of subjects in the observing groups. Data are shown for all 30 trials combined and for the first and second 15 trials. (Note: In the legend, the first 2 letters identify the goal of the model, the second 2 identify the goal of the observer. CT = Control Task; PS = Pattern Search.)	61
2.5	Mean percentage of correct responses to each category of prediction question for the sets of subjects in the observing groups.	62
3.1	Mean number of correct trials in the learning phase for the control task and dual goal groups. Data are shown for all 30 trials combined in the top chart and for the first and second 15 trials in the bottom chart. Experiment 1's data is included for the total phase scores.	83
3.2	Mean number of correct trials in the test phase for all the groups. Data are shown for all 30 trials combined in the top chart and for the first and second 15 trials in the bottom chart. Experiment 1's data is included for the total phase scores.	85

Figure number	Description	Page no.
3.3	Mean number of percentage responses to each category of prediction questions for each group including Experiment 1's data. The top chart shows data for the overall prediction questions score. The bottom chart shows the data for each question type (OC = Old-correct, OW = Old-wrong).	88
4.1	Mean number of correct trials in the learning phase for the specific and dual goal groups. Data are shown for all 30 trials combined and for the first and second 15 trials. Data were collapsed over secondary task as all groups concurrently verbalised during the learning phase.	113
4.2	Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials. Data were collapsed over secondary task as it had no effect on performance.	114
4.3	Mean percentage of correct responses to each category of prediction questions for each group. Data were collapsed over secondary task as it had no effect on performance.	116
5.1	Mean number of correct trials in the learning phase for the control task and dual goal groups. Data are shown for all 30 trials combined and for the first and second 15 trials.	136
5.2	Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.	137
5.3	Mean percentage of 'yes' answers for the different types of instances.	138
6.1	Mean number of correct trials in the learning phase for the control task subjects. Data are shown for all 30 trials combined and for the first and second 15 trials.	167
6.2	Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.	168
6.3	Mean percentage of correct responses to each category of prediction question for each group. Total prediction question scores are in the top graph, and individual types below.	170

Figure number	Description	Page no.
6.4	Mean number of correct trials in the transfer phase (and test phase for total scores) for each group. Data are shown for all 30 trials combined in the top graph and for the first and second 15 trials in the bottom. PIT = person interaction task, FT = factory task.	173
6.5	Mean number of correct trials in the learning phase for the control task subjects. Data are shown for all 30 trials combined and for the first and second 15 trials.	182
6.6	Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.	183
6.7	Mean percentage of correct responses to each category of prediction questions for each group.	185
6.8	Mean number of correct trials in the transfer phase (and test phase for total scores) for each group. Data are shown for all 30 trials combined in the top graph and for the first and second 15 trials in the bottom. PIT = person interaction task, FT = factory task.	188

Table of Tables

Table number	Description	Page no.
1.1	A Sequence of Possible Inputs and Outputs for the Person Interaction Task	23
1.2	Numbers of subjects giving each category of response to the rule description questions	33
2.1	Labeling of Observer groups	49
2.2	Ranking of answers to the rule description question.	58
2.3	Between groups comparisons for the prediction questions	63
2.4	Spearman Rank Correlations of observer's scores with their model's scores.	65
3.1	Ranking of answers to the rule description question section into each category for the two groups.	91
3.2	Comparisons with the non verbalising groups from chapter 1	92
4.1	Numbers of subjects giving each category of response to the rule description questions	118
5.1	Numbers of subjects giving each category of response to the rule description questions	140
6.1	Ranking of answers to the rule description questions into each category for the two groups.	172
6.2	Between groups comparisons for the total transfer phase score	172
6.3	Similarity ratings of the two tasks for each set of subjects	176
6.4	Ranking of answers to the rule description questions into each category for the two groups.	186
6.5	Similarity ratings of the two tasks for each set of subjects	190

DECLARATION

The research contained in this thesis was carried out by the author between October 1994 and June 1997 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

WORK FOR THIS THESIS HAS RESULTED IN THE FOLLOWING PUBLICATIONS:

Geddes, B.W., & Stevenson, R.J. (in press). Learning goals and the implicit and explicit learning distinction. *Quarterly Journal of Experimental Psychology.*

Geddes, B.W. & Stevenson, R.J. (1996). The relationship between goal specificity and whether learning is implicit or explicit. *Proceedings of the 9th Conference of the European Society for Cognitive Psychology,* page 87.

Stevenson, R.J., & Geddes B.W., (1997) Effects of Goal Specificity and Explanations on Instance Learning and Rule Learning. *Proceedings of the 19th Annual Conference of the Cognitive Science Society.* Stanford University, Palo Alto, Ca. LEA.

Introduction

This thesis explores the implicit-explicit learning distinction. The focus is upon one particular variable that has a dramatic impact on subjects' learning and important consequences for the distinction. The variable that is examined is the learning goal of a task. Essentially, this thesis shows that different learning goals can lead to different modes of learning - implicit / explicit (alternatively categorised as rule learning / instance learning). The effect of learning goal for the implicit-explicit learning distinction has previously only been briefly explored (Whittlesea and Dorken, 1993) and has not been examined at all in the context of the type of tasks used here.

The tasks used in this thesis fall into the category of dynamic system learning tasks. These are tasks that follow some underlying pattern or rule which guides the system's output. Subjects learning a dynamic system interact with the system by making inputs and receiving outputs. It is through these interactions that subjects learn something of the way the system works. The dynamic system used in this thesis is based on Berry and Broadbent's (1984) person interaction task. Berry & Broadbent showed that subjects are able to perform well at controlling the outputs, but are unable to describe the underlying pattern that the task follows. This result was taken to show that subjects learnt the task implicitly. Excluding Experiment 1 of this thesis (Geddes & Stevenson, in press), the only studies that have shown this task being learnt explicitly are when the task is altered so that the salience of the pattern is increased (e.g. Hayes and Broadbent, 1988; Berry and Broadbent, 1988; Dienes and Fahey; 1995). This variable of salience was thought to be the key factor in determining whether a task is learned implicitly or explicitly. The starting point of this thesis is to refute this notion by showing that there is a more fundamental determining variable - the learning goal of the task. The first experiment of this thesis shows that it is the specific control goal of the person interaction task that is responsible for it not being learned through explicit hypothesis testing and rule deduction. If the learning goal is altered to a non-specific, pattern search goal, then the same low salience pattern can be learned through explicit rule deduction resulting in the subjects being able to describe the underlying rule perfectly. Essentially, the remaining six experiments explore this finding and its theoretical implications.

Coverage of the relevant literature for the thesis appears separately in the introductions to each new experiment. The rest of this opening introduction, therefore, is a preview of the experimental work and of the theoretical implications of the 7 experiments presented in the thesis. Firstly though, to aid description of the experiments, the person interaction task and its measures of learning are described. Following this, the impetus behind the first experiment and the first experiment itself are introduced. The first experiment is the foundation paper upon which the other experiments are based. It is also the experiment where the theoretical model that is used throughout the thesis is presented. This is Klahr & Dunbar's (1988) dual space model of learning that neatly encompasses the findings of the thesis. Therefore, in the description of the first experiment this model is also briefly outlined. Then, for the rest of this opening introduction, the remaining experiments that explore the learning goal effect and the suitability of the dual space model are introduced.

The person interaction task is run on a computer and consists of subjects interacting with a computer person who is a construct of the programme operating the experiment. Both the subject and the computer person can be in one of twelve emotional states ranging along a scale from *Very Rude* to *Loving*. The computer person starts off being in one of the states (say *Polite)*. The subject responds to this with which ever state they choose (say *Affectionate*). The computer person then produces a new output (say *Loving*). The new output is governed by an underlying equation under which the dynamic system operates. Through making the inputs and receiving the outputs subjects should be able to learn something about the underlying equation.

There are a number of measures used to check subjects' learning of the dynamic system. One of the measures is the control performance measure which explores how proficient subjects are at making the system produce a specific output. In all the other studies that use the person interaction task (and also in some of the conditions used in the experiments in this thesis) this measure is also integral to the learning process. That is, in the initial learning trials the subjects are instructed to obtain a specific output and to maintain that output. In the computer person interaction task, subjects are typically given a specific control goal such as make the computer person be *Polite*. In a typical experiment, the number of correct learning trials (out of 30) is taken as a measure of learning. In the experiments reported in this thesis, the 30 learning trials are followed by 30 test trials in which subjects are instructed to make the computer person produce a novel specific output (e.g. *Very Friendly*). A second measure used to check subjects' learning is the rule description question. This is a question that asks subjects to explicitly state anything they can about the rule that the computer person operated under. Finally, a third measure that subjects are typically given is the prediction questions. For these, subjects are presented with sets of interactions such as "you were *Indifferent,* the computer person was *Polite*, you were then *Rude."* From these interactions subjects have to predict what output the computer person will be next.

A certain set of results from these measures was taken to indicate that subjects had learnt implicitly. If subjects were above chance at making the computer person produce the specific output then it was concluded that they had learnt something about the underlying system. However, if additionally subjects could not make predictions or answer the rule description question then it was concluded that they had no explicit knowledge of the task, therefore their good control performance was due to something they had learnt implicitly.

A key methodological enhancement used in this thesis improves the functionality of the prediction questions as an indicator of learning mode. In all other studies the sequence of interactions in the prediction questions from which subjects have to predict is created arbitrarily. That is, there is no attempt to distinguish between sequences that subjects have encountered before and novel sequences which subjects have not seen before. Nor is there a distinction between familiar sequences in which the subjects produced a correct response from the computer person and familiar sequences in which the subjects produced an incorrect response. For the experiments presented in this thesis the sets of interactions are created in a carefully controlled manner. Three types of prediction questions are used: Old-correct prediction questions that consist of familiar interactions in which the subjects had been successful at making the computer person produce the required output; Old-wrong prediction questions that consist of familiar trials in which the subjects had been unsuccessful at making the computer person produce the required output; and New prediction questions that consist of a sequence of unfamiliar trials.

Performance on these refined prediction questions allows a more detailed examination of the nature of subjects' learning. Success on Old prediction questions, but not New prediction questions would suggest that subjects' learning and post learning performance is purely due to some general memory effect. Success on only one type of Old prediction question would suggest that subjects learning and post task learning performance is due to some specific memory effect (e.g. memory for say, just trials subjects had performed correctly on). Success on all prediction questions including New questions suggests that subjects learning and post task learning performance is due the underlying rule of the task having been deduced. Without any memory for New prediction questions, subjects can only perform well on them by applying the rule that the system is based on. Thus by using these three kinds of prediction questions, the nature of the subject's learning can be determined.

The impetus behind the first experiment in this thesis came from a study by Owen and Sweller (1985). Owen and Sweller examined explicit learning in the classroom. They showed that a major influence on subjects' learning came from the learning goal. Basically, subjects given a non-specific goal gain more knowledge than subjects with a specific goal. School children were taught some of the basic principles of trigonometry. Some of these children (the specific goal group) had to practise on sets of problems that required specific solutions such as find angle A$\underline{B}$C. The other group (the non-specific goal group) had to practise on a set of problems that required no specific solutions but were just asked to find as many unknowns as possible. The problems were arranged so that both groups were calculating the same number of unknowns. The non-specific group clearly learnt better than the specific group: They made fewer mistakes during a test phase and their performance was superior to that of the specific group on novel problems.

The idea behind the first experiment of this thesis was that this goal effect may be able to explain the apparent implicit learning of low salient dynamic systems control tasks. In all other studies of the person interaction task subjects have always had a specific learning goal. It is quite possible that explicit hypothesis testing and rule deduction are prevented by this specific control goal. To test this hypothesis Experiment 1 uses Berry & Broadbent's (1984) Clegg version of the person interaction task. (The computer person is called Clegg, and Clegg follows one particular underlying rule.) To explore the possibility of a learning goal effect, for the learning phase of the task, three different groups of subjects are used, each one with a different learning goal. The goals are; a specific control goal like that used by Berry & Broadbent; a non-specific pattern search goal where, rather than asking the subjects to make the system produce a specific output, the goal is simply to try to determine the pattern that Clegg is following; and finally, a dual goal that is included to complete the design and is a combination of the other two goals.

The other major difference in Experiment 1 from Berry & Broadbent's study is the new way the prediction questions are created (as described above). In relation to Experiment 1, there are two reasons for creating the prediction questions in the new fashion. The first is that the non-specific goal group in Experiment 1 should be learning by explicitly deducing the underlying rule that Clegg is following. Therefore, a distinct set of New prediction questions is needed to test this proposition. The problem with the arbitrary way of creating prediction questions as used in earlier studies, is that they may consist solely of Old sequences and therefore good performance could be explained as some sort of memory effect. The New questions consist of a novel situation for subjects to predict from, therefore, the only way to make a correct prediction is if subjects have learned the underlying rule. Any group that is supposedly learning rules should be able to predict successfully on New prediction questions. The other reason to create the prediction questions in the new manner is for the inclusion of a specific set of Old prediction questions. One way to look at the implicit / explicit learning distinction is by thinking about it as an instance learning / rule learning distinction. Very briefly, the idea is, that implicit learning is actually a memory effect and therefore can be described as subjects memorising instances. This distinction ignores controversial issues about the state of awareness during learning episodes. If implicit learning can be termed as instance learning then supposed implicit learners should be able to perform well on Old instances and not New instances. There is also the possibility that implicit (or instance) learners are only recording certain instances, such as only the instances that they achieve their learning goal on. The discrete set of Old-wrong and Old-correct prediction questions used throughout the experiments in the thesis allow these issues to be examined.

The learning goal effect is clearly shown in Experiment 1. Subjects with a non-specific pattern search goal are able to learn explicit rules. Apart from anything else, this is shown by the fact that these subjects can clearly state the rule the system follows. The results for the specific goal group replicate Berry and Broadbent's results. Subjects show good control performance, but are not able to state the rule that Clegg followed. Additional to Berry and Broadbent's results, the prediction questions show that the specific goal subjects' learning is in line with an instance learning model where all instances are recorded. There is also a surprise result from the dual goal group. These also show a pattern of learning that can be described as instance learning, not rule learning. However these subjects appear to be only remembering instances they perform correctly on. Unlike the specific goal group who perform well on both correct and incorrect Old prediction questions, dual goal subjects are better on Old-correct questions than on Old-wrong questions.

Klahr and Dunbar's (1988) dual space model of learning is used to explain the learning goal effect. Briefly, this model suggests that learning can operate in two spaces - rule space and instance space. It can explain the learning goal effect shown in Experiment 1 because it suggests that the specific control task of both the specific and dual goal subjects encourages subjects to explore only instance space and prevents them from exploring rule space. The non-specific pattern search goal however encourages the exploration of both rule space and instance space - rule space to hypothesise about the underlying rule and instance space to test the hypotheses. Notably the dual space model suggests that the differences in learning with the different learning goals occur as a result of a *direct* cognitive effect of the learning goal.

Experiment 2 looks at whether the learning goal effect could be caused by an *indirect* effect of learning goal. One of the differences between a specific and non-specific goal is that for the specific goal the range of interactions of the dynamic system that subjects are exposed to should be much narrower. As subjects are trying to achieve their specific goal they should hopefully only be getting outputs from the system that are on the goal or near to it. Due to the nature of the underlying equation, to keep Clegg on the specific required output, the subjects' input should also be very near to the output. Therefore, the set of interactions that subjects are exposed to should represent only a very narrow range of the interactions of that the system allows. A recent study by Buchner, Funke, & Berry (1995), provided evidence about the range of interactions subjects encounter affecting their post task performance. In Experiment 1 it is possible that non-specific goal subjects are explicitly learning rules because they are exposed to a wider range of interactions than the specific goal subjects. In effect, this may mean that the pattern is made more salient. This would suggest that the learning goal has an indirect effect by simply increasing the salience of the pattern, not a direct cognitive effect by encouraging subjects to explore rule space. To test this, observer subjects with a non-specific learning goal learn by observing a range of interactions that were produced by model subjects who had a specific control goal. The results show that the observer subjects still learn explicit rules suggesting that the learning goal effect is due to a *direct* cognitive influence of the goals.

Experiment 3 explores what the direct cognitive effect of the learning goals might be. In line with the dual space model, the experiment tests the proposition that a non-specific pattern search goal encourages subjects to hypothesise and explore mainly rule space whereas the specific control goal encourages subjects to explore only instance space. This is done by giving the different learning goal subjects a secondary task which is either compatible or incompatible with their proposed cognitive processes. The secondary tasks ask subjects to either describe aloud what they are doing or to explain what they are doing. The explain secondary task also allows the experiment to test the effects of explanations on learning, since the positive effect of self explanations has been demonstrated by a number of studies (e.g. for improvement in; physics learning - Chi, Bassok, Lewis, Reimann, & Glaser, 1989; computer programming - Pirolli & Bielaczyc, 1989; Pirolli & Recker, 1994; general text comprehension - Chi, de Leeuw, Chiu, & LaVancher, 1994). However, the effect of self explanation has not been specifically tested on a dynamic systems task. The proposed compatible conditions of learning goal and secondary task proved to be just that, confirming the interpretation of the cognitive processes that different learning goals induce. The results therefore, further endorsed the dual space model as a theoretical umbrella under which to explain the learning goal effect. Additionally, the results also demonstrate the positive effect of self explanations on learning a dynamic systems task.

Experiment 4 is designed to explore the effect of the dual learning goal. The dual goal subjects have a combination of the control goal and pattern search goal. Experiment 1 showed that these subjects appear to be performing instance learning but by focusing only on instances in which they achieve the required output. The hall mark of their learning is that they perform better on Old-correct prediction questions than on Old-wrong prediction questions. In Experiment 1, it is suggested that these subjects may be learning purely implicitly as their dual goal creates a cognitive load that is too heavy for explicit processes. To test this proposal the three learning goal groups are given a secondary task of concurrent random number generation to occupy the central executive of working memory and prevent explicit learning processes. The results show that for all three learning goals, subjects perform comparably to the dual goal subjects in Experiment 1. All subjects make correct predictions from Old-correct prediction questions only. These results are taken as confirmation that dual goal subjects, learning the Clegg version of the person interaction task, are learning purely implicitly.

Experiment 5 is designed to explore the effect of the learning goals and resulting modes of learning on memory. The experiment allows the exploration of whether explicit rule learners are exclusively learning rules or learning instances as well. The work of Nosofsky, Clark & Shin (1989), suggests that rule learners learn only rules not instances. Experiment 5 tests this proposition by giving subjects a recognition test instead of prediction questions. The subjects are asked to decide whether or not they have seen sequences of 2 trials before. The experiment also allows the examination of the suitability of the model that is used to explain the two different forms of instance learners' results. In Experiment 1, it is suggested that instance learners construct a look-up table based on their learning experiences (Dienes and Fahey, 1995). Then, for any post learning task that subjects perform, they refer to the look-up table for the appropriate action or solution to the new task. Such a model explains the prediction question performance in Experiment 1, since instance learners were good at Old prediction questions that were entered into their look-up table but were poor at New prediction questions that obviously had no entry in the look-up table. Notably, the look-up table, which is supposedly a reflection of memory, suggests that dual goal subjects should have no memory for incorrect instances. The results confirm the predictions for the control goal subjects. Their memories consist of an even balance of Old-wrong and Old-correct instances and therefore their look-up tables can be considered to represent all of what is in their memories. The results only partially confirm the predictions for the dual goal subjects. They have better memories for the Old-correct instances than for the Old-wrong instances. However their memory for Old-wrong instances is also quite good. This suggests that dual goal subjects' look-up table only consists of part of what is in memory. The results for the explicit rule learners suggest that they do have some memory for instances; however amongst other things, this may be due to the late placing of the memory test.

Finally, the last two experiments, 6a and 6b, explore whether or not the learning induced by different goals leads to transfer to a novel but structurally identical task. The novel task used is Berry & Broadbent's (1984) sugar production task. The inputs and outputs in this transfer task follow an identical pattern to those of the person-computer interaction learning task. The transfer abilities of both rule and instance learners of the same dynamic systems task has never been explored. The look-up table model used to describe instance learners' abilities would suggest that instance learners should not be able to transfer to a novel task which, though structurally identical, is perceptually different. Explicit rule learners on the other hand should be able to transfer what they have learned to such a novel task (e.g. Gick and Holyoak, 1983; Owen and Sweller, 1985; Sweller, Mawer and Ward, 1983; Vollmeyer, Burns and Holyoak, 1996). Indeed, it can almost be seen as the ultimate test of explicit learning. The results of Experiment 6a confirmed the prediction about the instance learners, however only partially confirmed the prediction about the explicit rule learners. A close analysis of the sugar production task suggested that it was actually more complex than the person interaction task. Therefore, Experiment 6b was designed to test transfer between two tasks that should be of comparable complexity. The novel task this time was a simplified version of the sugar production task. This time the predicted results for the instance and rule learners held true.

Chapter I

Experiment 1: Learning Goals And Implicit vs Explicit Learning

An assumption common to several current theories of human learning is that learning proceeds by means of two separate systems (e.g. Berry and Broadbent, 1984; 1988; Broadbent, Fitzgerald and Broadbent, 1986; Reber, 1989; Shanks and St. John, 1994). Some proponents of this view contend that the two systems give rise to qualitatively different forms of knowledge (e.g. Broadbent et al, 1986; Reber, 1989) while others contend that there is a single knowledge base upon which two distinct learning processes can operate (e.g. Shanks and St. John, 1994). Some investigators additionally propose that the systems involve two different states of awareness at learning (Reber, 1989; Hayes and Broadbent, 1988).

Support for this notion of separable learning systems can be found in studies of implicit learning, which show dissociations in performance across different experimental tasks (e.g. Berry and Broadbent, 1984; Hayes and Broadbent, 1988; Reber, 1967). The term "implicit learning" was first coined by Reber (1967) who defined it as the acquisition of complex abstract knowledge that takes place without the learner's awareness that he or she is learning. Knowledge acquired during implicit learning is not reportable. Explicit learning, in contrast, proceeds with the subject's awareness of what is being learned, and the knowledge that is acquired is verbally reportable. Some research suggests that implicit and explicit learning can be functionally dissociated (e.g. Berry and Broadbent, 1984; Hayes and Broadbent, 1988; Reber, 1967, 1989). However, other research suggests that the two learning systems can be used in combination (Buchner, Funke and Berry, 1995; Dienes and Fahey, 1995; Reber, Kassin, Lewis and Cantor, 1980).

Recently, however, the distinction between implicit and explicit learning has been contested (Shanks and St. John, 1994). Shanks and St. John argue that there has been no clear cut evidence for a dissociation between implicit and explicit learning, evidence that they claim is needed to prove the existence of the distinction. In the absence of such evidence, Shanks and St. John favour the view that while there *are* two dissociable learning systems, they are both explicit and do not involve unconscious processes. There are a number of reasons for being cautious about the conclusions of Shanks and St. John. For example, they propose two criteria that should both be met before a dissociation between what is learned and what can be reported is assumed[1]. While a number of commentators endorse these proposed criteria (e.g. Berry, 1994; Perruchet and Galliego, 1994) others regard them as too stringent (e.g. Dienes and Perner, 1994; Holyoak and Gattis, 1994). Also, it is likely that there is a continuum of processing from implicit to explicit rather than a clear dichotomy (Cleeremans, 1994). (See also the other commentaries on Shanks and St. John's (1994) target article.)

There is however, considerably more agreement over the distinctive *nature* of the two learning processes, regardless of whether they are implicit or explicit. Many investigators propose that what is commonly regarded as implicit learning comes about through the memorisation of instances encountered during learning (e.g. Broadbent et al, 1986; Dienes and Fahey, 1995; Shanks and St. John, 1994), and that this process is best modelled by a PDP connectionist network (e.g. Cleeremans, 1993). By contrast, what is commonly regarded as explicit learning comes about through hypothesis generation and testing (e.g. Simon and Lea, 1974; Klahr and Dunbar, 1988; Shanks and St. John, 1994) and may therefore be best modelled by a serial symbolic network.

A number of researchers have cautioned against the use of strict dichotomies, such as the one between two learning systems. Green and Shanks (1993), for instance, point out that there is evidence for two (possibly three) memory subsystems underlying instance learning: perceptual, motor and (possibly) conceptual subsystems. Cleeremans (1994) comments that a simple recurrent network, which learns by processing instances of a finite state grammar, can develop internal representations ranging from the raw storage of instances to fully abstract representations. Whittlesea and Dorken's (1993) evidence, to be discussed later, supports this view. Finally, researchers have shown that in implicit learning tasks, subjects may adopt a range of explicit learning strategies in addition to learning instances and that which kind of learning predominates depends on the way the learning task is presented (e.g. Buchner et al, 1995), the salience of the rule (e.g. Berry and Broadbent, 1988), and the instructions to the subjects (e.g. Reber, Kassin, Lewis and Cantor, 1980).

---

[1] Shanks and St. John's information criterion states that the measure of awareness must test the same information as the measure of performance, while the sensitivity criterion states that the measure of awareness must be sensitive to all conscious knowledge.

In this chapter, the inter-relationship between these two types of learning is further investigated by examining the ways in which learning goals affect the extent to which each kind of learning predominates. In particular, the idea that a specific learning goal encourages instance learning while a non-specific learning goal encourages explicit hypothesis generation and testing is explored. The two types of learning are referred to as instance learning and rule learning rather than using the implicit/explicit distinction. However, the term "implicit" is retained when referring to tasks commonly discussed in the implicit learning literature and the term "explicit" when referring to tasks commonly discussed in the explicit learning literature (that is, the literature on problem solving and hypothesis testing). In the experiment presented in this chapter, one of Berry and Broadbent's (1984) dynamic control tasks is used. This introduction continues, therefore, with a description of this task and the relevant research findings. A review of evidence for the importance of learning goals in implicit tasks on the one hand and explicit tasks on the other is then presented. This evidence forms the background to the hypothesis that specificity of the learning goal determines whether instance learning or rule learning predominates in a non-salient task previously found to be impervious to rule learning.

Berry and Broadbent's influential (1984) study showed an apparent dissociation between learning and awareness. One of their tasks (the one used in the present study) required subjects to interact with a 'computer person' called Clegg and try to get him to become and stay *Very Friendly*. Clegg initiated the interaction by displaying one of twelve attitudes (e.g. *Polite, Very Friendly, Loving*) on the computer screen, after which the subject had to respond by typing in another attitude. The attitudes reflected an intimacy scale from very low to high and Clegg responded to the subject's choice of attitude in an overreactive manner. If Clegg had displayed *Polite*, and the subject responded with the attitude, *friendly*, then Clegg would retaliate with the attitude *Loving*. Clegg's attitude on each trial was a simple numerical function of the subject's response on that trial and Clegg's previous output. The subjects successfully learned to carry out this task. However, when questioned about the experiment afterwards, they were unable to describe what they were doing or what the underlying rule was.

However, there was some evidence for rule learning in a second experiment. When subjects were given instructions that were, in essence, a verbal description of the equation governing the computer's attitude, the subjects outscored those who had no instruction and gave evidence of rule learning. In subsequent work, further attempts were made to facilitate rule learning. It was found that when the rule underlying Clegg's behaviour was simplified so that Clegg's response on a trial was a function of only the subject's response, subjects did show explicit knowledge of the rule when they were questioned at the end of learning (Hayes and Broadbent, 1988; Berry and Broadbent, 1988; Dienes and Fahey; 1995). Thus, when the rule is described to the subjects or when the rule is made more salient, rule learning can take place. However, under normal circumstances, the typical finding is that subjects learn to control the computer but give no evidence that they have learned the underlying rule.

A study by Stanley, Mathews, Buss & Kotler-Cope (1989) further highlights the difficulty of rule learning in this task. Stanley et al used the same person interaction task as Berry & Broadbent (1984) but devised a different measure of verbalisable knowledge. At the end of every 10 trials subjects had to verbalise information that they had acquired from controlling the computer person. Subjects were told that this information was intended to aid yoked subjects in controlling the computer person. It was found that this information only aided yoked subjects' ability to learn the task when the information came from subjects who had completed over 570 trials of controlling the computer person.

Arguably, strong evidence for rule learning during the person interaction task with non-salient rules has never been shown. Complete evidence should include a number of results: superior control performance in a post-learning test phase; the ability to predict, on the basis of the learned rule, the computer's next response when presented with a sequence of inputs and outputs not encountered before; the ability to state explicitly the rule underlying the computer's behaviour; and significant positive correlations between measures of control performance and measures of verbalisable knowledge. While Dienes and Fahey (1995) found good performance on familiar prediction questions, they found poor performance on novel prediction questions. Also, Berry and Broadbent (1984, 1988) found no improvement on a rule description question asking for explicit knowledge of the rule and no significant positive correlations between control performance and questionnaire performance. They did not test for control performance in a test phase. All in all, therefore, strong evidence of explicit learning has not yet been observed.

However, the methods used to investigate learning in the person computer interaction task all give subjects the same specific learning goal - to make the computer person produce and maintain a specified output. The hypothesis explored here is that this goal specificity is responsible for the low level of rule learning that has been observed. This hypothesis is based on the evidence suggesting that the learning goal can have a profound effect on the kind of learning engaged in by the subjects, whether it be instance learning (Whittlesea and Dorken, 1993) or rule learning (Owen and Sweller, 1985; Sweller, 1988; Vollmeyer and Burns, 1995; Vollmeyer, Burns and Holyoak, 1996).

As regards instance learning, Whittlesea and Dorken conducted a series of studies on the learning of artificial grammars showing that what is learned about an instance depends on the kind of processing the subject engages in during learning. According to their "episodic-processing" account, what gets represented is the way the instances are processed and not the instances themselves. That is, Whittlesea and Dorken argue that what is learned is a function of the processing guided by the purpose of the task, and is not simply a function of the structure of the items. The episodic-processing account also says that either specific or general knowledge can develop during instance learning depending on the particular conditions of the experiment.

For example, in one experiment (Experiment 5a), Whittlesea and Dorken presented subjects with exemplars of an artificial grammar in what was announced as a number learning experiment. Subjects were required to spell the letter strings, which were presented as distracters to prevent rehearsal of the numbers. At test, subjects were asked to say whether the test items were grammatical or ungrammatical. When the test items contained the same letter set as the ones presented during learning, grammaticality judgements were better than chance. But when the test items contained novel letters, not seen before, then the judgements were at chance level. Whittlesea and Dorken concluded that subjects had learned about the surface structures of the items but not their deep structures. However, in another experiment (Experiment 5b), subjects were asked to judge whether each letter of a training string was repeated elsewhere in the string. This task engaged processing of the repetition patterns that were common across items so that deep structure knowledge could be abstracted from the items. In these circumstances, subjects performed the grammaticality judgements above chance on test items containing the novel letter set as well as those containing the familiar letter set. Thus, the level of abstractness of the acquired knowledge depends on the processing induced by the learning task.

As regards rule learning, there are two research areas in which such learning is emphasised. One is the area of problem solving (Newell and Simon, 1972); the other is the area of the hypothesis testing (e.g. Simon and Lea, 1974; Klahr and Dunbar, 1988). In both these areas, researchers assume that the learning in question is explicit. Researchers in the area of problem solving assume that people with little knowledge of a domain discover solutions to novel problems in that domain by using general problem solving methods, often referred to as "weak methods" because they require little background knowledge. Means-ends analysis is one such method. It involves breaking each problem down into sub-problems and solving each sub-problem using a difference reduction strategy. The general feature is that the current state is matched against the goal state and if there are differences between the two, sub-goals are established to eliminate the differences. As each sub-goal is met, the procedure is repeated until the current state matches the goal state. One of the most influential cognitive theories of learning is rooted in the problem solving tradition (Anderson, 1987). According to Anderson, the acquisition of a cognitive skill comes about when declarative knowledge, in the form of instructions, is converted into procedural knowledge in the form of production rules. In the theory, the process of conversion is achieved using general purpose weak methods, such as means-ends analysis, to convert declarative knowledge into domain-specific productions via a mechanism of compilation.

However, the generality of this learning model has been challenged (Owen and Sweller, 1985; Sweller, 1988; Vollmeyer and Burns, 1995; Vollmeyer, Burns and Holyoak, 1996). Sweller argues that use of means-ends analysis and similar problem solving methods prevent the acquisition of abstract (conceptual) knowledge. According to Sweller and to Vollmeyer et al, means-ends analysis can be applied to a well-defined problem with a specific goal. However, the method does not result in rule induction because it only yields a solution path to the specific goal. On the other hand, alternative learning strategies, such as hypothesis testing, do result in rule induction because they involve a wider search of the problem space. Sweller and Vollmeyer et al argue further that a non-specific goal will facilitate such a wide search and so foster rule learning.

In the area of hypothesis testing, "dual space" models provide a mechanism whereby rule learning and problem solving can be explained within the same framework[2]. Simon and Lea proposed that the problem space is separated into two spaces: a rule space and an instance space. People search instance space when seeking the solution to a specific goal; that is, when they are problem solving. In problem solving, legal operators are applied to a state to generate a new state. However, when hypothesis testing, people search both rule space and instance space. Explicit rules or hypotheses are generated in rule space, and these rules are tested by experiments that generate states in instance space. Thus, according to Simon and Lea, problem solving takes place in instance space while hypothesis testing takes place in both spaces. Klahr and Dunbar (1988) have adopted a similar dual space model, consisting of hypothesis space (comparable to Simon and Lea's rule space) and experiment space (comparable to Simon and Lea's instance space). As in Simon and Lea's model, hypotheses are generated and modified in hypothesis space and tested in experiment space.

A number of studies support these ideas. First, Vollmeyer et al (1996) showed that when given a non-specific goal for learning a complex control task, subjects used explicit hypothesis testing strategies. Second, Sweller and his co-workers have shown that people with a non-specific goal gained more knowledge about a task than did people with a specific goal (e.g. Mawer and Sweller, 1982; Owen and Sweller, 1985; Sweller, 1988). For example, Owen & Sweller (1985) taught school children some of the basic principles of trigonometry (cosine, sine, tangent). They then asked separate groups of these children to practise on sets of problems with goals differing

---

[2] It is possible that instance learning can also be integrated into the framework, with instance learning occurring in instance space as well as problem solving. Such an integration would provide a framework for understanding how it is that people frequently show a mixture of instance learning and rule learning in the same task.

in specificity. One group (the specific group) had to solve problems with typical, specific goals (e.g. find the length AB on triangle ABC and also find angle A). The other group (the non-specific group) were given a similar triangle and simply told to find as many unknowns as possible. The problems were arranged so that both groups were calculating the same number of sides and angles. The non-specific group clearly learnt better than the specific group: They made fewer mistakes during a test phase and their performance was superior to that of the specific group on novel problems.

Vollmeyer and Burns (1995) and Vollmeyer et al (1996) found comparable results with their complex control task. Vollmeyer et al asked subjects to learn a system in which four different water quality factors (temperature, salt, oxygen and current) had varying degrees of influence on the numbers of four different sea animals (prawns, sea bass, lobster and crabs). The subjects' task was to determine the links between water quality factors and animals, together with the strength of these links. While all the subjects were told to explore the system so as to learn as much as possible, half were also told the specific goal on which they would be tested after learning. The results showed clear effects of goal specificity. In contrast to the specific goal group, non-specific goal subjects gained more knowledge of the system (measured by diagrams drawn at the end of learning), and transferred what they had learned to a novel specific goal.

As has been seen, means-ends analysis is said to occur in instance space (Simon and Lea, 1974). With sufficient practise, it results in the acquisition of production rules that encode the subject's current state with the following goal (or sub-goal) state. It can, therefore, be seen as an example of instance learning. Memory array models have been used to explain instance learning (e.g. Medin and Schafer, 1987); in these models, learning of production rules, such as those described above, result in a look-up table in which individual instances are stored in memory (e.g. Mathews, 1991; Perruchet, 1992; Dienes & Fahey, 1995; Buchner, Funke, & Berry, 1995). The main aim of the study reported here is to test the proposition that a specific goal leads to instance learning in a dynamic control task while a non-specific goal leads to rule learning. In the light of the results reviewed above, the non-specific goal subjects would be expected to produce superior learning compared to specific goal subjects.

More specifically, in relation to the measures used in implicit learning tasks, non-specific goal subjects should be superior at reaching a specific goal in a post-learning test phase, and at predicting the computer's next response in a novel situation (a sequence of three possible input and output states). As their learning is expected to be generally superior it would also be expected that these subjects show superior prediction performance in familiar situations. Finally, the non-specific goal subjects should be able to describe the rule underlying the computer's behaviour when asked to do so, and there should be positive correlations between control performance and answers to the prediction questions. The specific goal subjects should transfer their learning less readily to a novel specific goal at the test phase, and they should also be able to predict the computer's next response in familiar situations but not in novel situations. Some versions of the memory array model suggest that only instances where subjects reach their goal are recorded in the look-up table (Dienes & Fahey, 1995; Marescaux, Luc, & Karnas, 1989). Therefore it would also be expected that specific goal subjects will perform better on familiar situations that they have performed correctly than on ones where they were incorrect. Finally, specific goal subjects should be unable to describe the rule underlying the computer's behaviour and there should be no correlation between control performance and answers to the prediction questions.

Also included in this experiment is a group of subjects who were given both goals. Having two goals should cause difficulty for the subjects, since the goals make incompatible demands. The specific goal will direct subjects to search only the instance space while the non-specific goal will direct them to search both instance and rule spaces. Indeed, Berry & Broadbent (1988) found that learning of a non-salient dynamic control task was considerably impaired when subjects had to look for an underlying pattern (a non-specific goal) as well as reach a specified goal state (a specific goal). Thus a similarly impaired performance would be expected with a dual goal. What is of interest though, is whether or not there can be evidence for any systematic learning at all when subjects have to meet two goals, and if so, what kind of learning is preserved, instance learning or rule learning.

METHOD

Subjects: The 72 volunteer subjects were Durham University graduate and undergraduate students, aged between 18 and 24.

Design: Subjects were randomly allocated to one of three goal groups (a specific goal group, a non-specific goal group and a dual goal group). All subjects were required to complete 30 learning and 30 test trials. The goal groups were defined by the nature of the goal in the learning trials. The specific goal group was given a specific goal, the non-specific goal group was given a non-specific goal and the dual goal group was given both the specific and the non-specific goals. In the test trials, all subjects were given a new specific goal. There were two specific goals: to make Clegg *Polite* and to make Clegg *Very Friendly.* Half the subjects in the specific and dual goal groups made Clegg *Polite* in the learning trials and *Very Friendly* in the test trials. The order of these two goals was reversed for the remaining subjects. After the test trials, all subjects were given an unexpected questionnaire, consisting of 15 prediction questions followed by two rule description questions, which probed explicit knowledge of the rule. There were 3 types of prediction question - 5 based on old correct trials, 5 on old incorrect trials and 5 on new trials. These 15 questions were displayed in a random order. The order of the two rule description questions was counterbalanced across subjects .

The task : Subjects were told that they would be meeting a computer person named Clegg and would communicate to Clegg through the screen and keyboard. Clegg would express his attitude towards them by displaying one of twelve descriptions (*Very Rude, Rude, Very Cool, Cool, Indifferent, Polite, Very Polite, Friendly, Very Friendly, Affectionate, Very Affectionate, Loving*). Following this, subjects responded to Clegg by choosing one of the above descriptions. This was done by typing in the first letter or letters of that description (e.g. VP for *Very Polite*). Once subjects had responded to Clegg he would display his new attitude (produced by the equation described below). It would then be the subject's turn to enter their next attitude, and so on. The list of possible responses was displayed on a piece of paper attached to the bottom of the screen for permanent reference.

In addition to the above instructions, each group of subjects was given specific instructions concerning their learning goal. Half the subjects in the specific goal group were told "Your aim is to shift Clegg to the *Polite* level and maintain him at that level". The remaining half were told to "shift Clegg to the *Very Friendly* level and maintain him at that level". The non-specific goal group was told "Your aim is to establish under what pattern Clegg is reacting". Half the dual goal group were told "Your aim is to shift Clegg to the *Polite* level and maintain him at that level. You should also try to establish under what pattern Clegg is reacting". The remaining half were told to "shift Clegg to the *Very Friendly* level and maintain him at that level" before being told to try to establish Clegg's pattern of responding. To remind subjects of their respective goals, the goal of their task was permanently displayed on a piece of paper attached to the bottom of the screen. The rest of the experiment was identical for all three groups, except that half the subjects had to make Clegg *Very Friendly* in the test trial and the other half had to make him *Polite*.

On each trial Clegg's and the subject's responses were displayed on the screen. These scrolled up the screen so that it was possible to see the previous six trials on the screen at any one time. The equation relating Clegg's responses to those of the subject's was identical to the non-salient rule used by Berry & Broadbent (1984) and Stanley et al (1989). The descriptions were given a value from 1 (*Very Rude*) to 12 (*Loving*). Clegg's response was determined by the equation :

$$CNR = (2 \times SOR) - COR + Z,$$

where CNR = Clegg's new response, SOR = subject's old response, COR = Clegg's old response and Z = a random number with the value of -1, 0 or +1. If Clegg's calculated new response was higher than the top response of the 12 point scale then his response reverted to the top response. If Clegg's calculated new response was lower than the bottom response of the 12 point scale then his response reverted to the bottom response. Table 1.1 is a display of typical inputs and outputs that the equation could generate. The random element in the equation ensures that subjects must exercise continuous control over the computer person. It also means that there is no unique input associated with any one output. If subjects reached their target output then simply re-entering the same input is unlikely to keep them on target (Berry & Broadbent, 1984). The optimum strategy for the subjects with specific goals to move Clegg successfully onto target would be to go half way between Clegg's old response and the target value on the behaviour scale. To allow for the random element in the equation producing Clegg's response, the responses of subjects in the specific and dual goal groups, were scored as correct if they were either on the target or one response either side of the target. That is, a response from Clegg of *Indifferent, Polite,* or *Very Polite* was scored as correct when the goal was to make Clegg Polite, while a response of *Friendly, Very Friendly,* or *Affectionate* was scored as correct when the goal was to make Clegg *Very Friendly.*

Table 1.1
A Sequence of Possible Inputs and Outputs for the Person Interaction Task

Subject's Response **(Input)**	*Clegg's Response* **(Output)**
	Polite
Very Polite	*Friendly*
Very Friendly	*Affectionate*
Polite	*Very Cool*
Friendly	*Loving*
Polite	*Very Rude*
Indifferent	*Very Friendly*
Very Friendly	*Friendly*

The test trials were identical to the learning trials for the specific goal group except that the goal was changed. Half the subjects in each group had to make Clegg *Polite* and maintain him at that level; the other half had to make and keep him *Very Friendly.* As was the case in the learning trials, a response either on the target or one step either side of the target was scored as correct, to allow for the random element in the equation.

The Questionnaire

Prediction questions: There were 15 prediction questions. Each question took the following form: First a typical trial situation was presented. The subject's and Clegg's behaviour was displayed on the screen, below this the subject's new behaviour was displayed - e.g. You were *Very Cool,* Clegg then was *Very Rude,* You were then *Polite.* Subjects then had to predict what Clegg's response would be. The trial situations were generated in three different ways:

A) 'New' situations; Each situation was generated randomly from a list of all possible trial situations that the subject had not encountered during either the learning trials or the testing trials.

B) 'Old-wrong' situations; Each situation was randomly selected from all the trials the subject had got wrong during the test phase.

C) 'Old-correct' situations; Each situation was randomly selected from all the trials the subject had got correct during the test phase.

Five questions of each type were produced. To produce five Old-wrong and five Old-correct questions meant that the subject must get at least five wrong or five correct respectively during the test trials. The program controlling the experiment allowed for the possibility of this not occurring and would have substituted any uncreated questions with New questions.

Rule description questions: There were two rule description questions in response to which subjects were asked to write freely, not worrying about wording or grammaticality. The two questions were:

A) "How did you get Clegg to behave as you wanted him to?" This question was designed to be sensitive to verbal knowledge of a procedural nature that may have been acquired during the experiment.

B) "Could you try to describe what sort of pattern you thought Clegg was using to respond to your behaviour?" This question was designed to be sensitive to knowledge of a more declarative nature that may have been acquired during the experiment.

Previously Berry & Broadbent (1984) had included a rule description question, similar to question (A), that was only guided towards tapping verbal knowledge of the procedural nature of the task. It was hoped that by having two different forms of the question, subjects would also have a chance to express verbally their knowledge connected to the closely related declarative side to the task - the underlying pattern of Clegg's behaviour.

Procedure: Subjects were randomly allocated to one of the three goal groups. As mentioned above, the three groups received identical initial instructions apart from one sentence. This sentence dictated the aim of that particular group for the learning trials. Apart from this, the remainder of the experiment was identical for all groups:

The instructions explaining the nature and aim of the subjects' initial learning task were presented first. These were followed by the learning trials. On completion of this phase, subjects from the three groups received instructions describing their new goal for the test trials and then the test trials started. Clegg initiated both learning and test trials by displaying one of the three adjectives centred on *Polite*. Following the test trials the subjects were presented with instructions for the prediction questions. These instructions described the nature of the questions and gave an example of a situation from which the subjects would have to make a prediction. The instructions also explained that each question was unrelated to the previous one. After completing the prediction questions subjects were given a pen and paper and were asked to answer the two rule description questions appearing on the paper.

Throughout the experiment, all instructions appeared on the computer screen but were also read out to the subjects. The experimenter stayed with the subject throughout the experiment in order to answer any arising questions.

RESULTS

Preliminary analyses indicated that there was no significant effects of goal order (*Polite* then *Very Friendly* vs. *Very Friendly* then *Polite*) on any of the measures reported below. Nor were there any significant interactions between goal order and any of the other variables. Consequently, to maintain clarity, the data were collapsed over the two orders in all the analyses.

Learning Trials

As mentioned before, learning trials were scored as correct for the specific and dual goal groups if they got a response from Clegg of *Indifferent, Polite* or *Very Polite* when they had the *Polite* goal and if they got a response of *Friendly, Very Friendly* or *Affectionate* when they had the *Very Friendly* goal. This scoring takes into account the random element of the equation producing Clegg's behaviour. Due to the lack of a specific goal for the non-specific goal group during their learning phase, no measure could be made for their performance. So only the learning of the specific and dual goal groups could be assessed, since only these two groups had a specific learning goal. The mean number of correct learning trials for these two groups are shown in Figure 1.1. In the Figure, data are shown for all 30 learning trials combined and for each half of the learning trials.

The data in Figure 1.1 were analysed using a 2 (learning goal: specific vs dual) by 2 (trial block: first 15 trials vs last 15 trials) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(1,44) = 24.79$, $p < 0.001$: performance was better for the specific goal group than for the dual goal group. There was also a main effect of trial block, $F(1,44) = 4.41$, $p = 0.042$: there were more trials correct in the second 15 trials than in the first 15. There were no significant interactions. (See Appendix 1 for the ANOVA tables and full sets of t-tests for this experiment, pg.228)

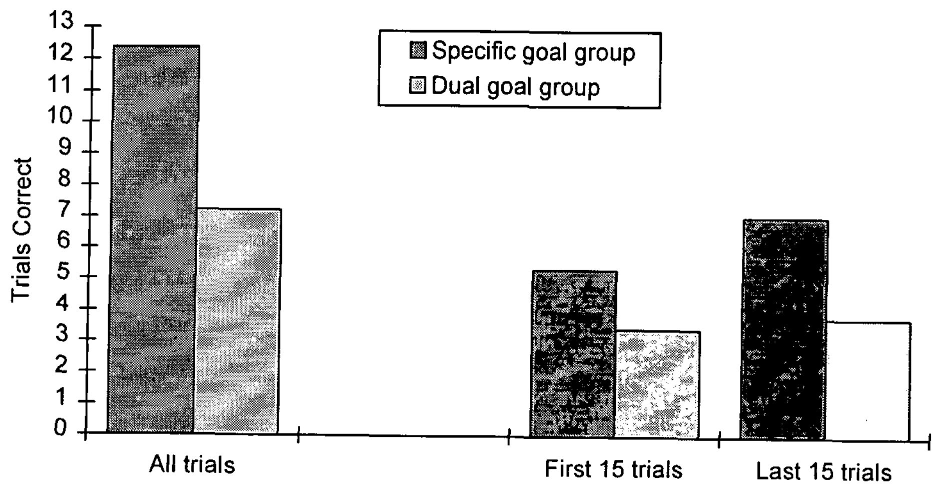


Figure 1.1: Mean number of correct trials in the learning phase for the specific and dual goal groups. Data are shown for all 30 trials combined and for the first and second 15 trials.

Test Trials

For all three goal groups, correct trials were identified in the same way as for the learning trials. Figure 1.2 shows the mean number of correct test trials for each group for the entire test phase and each half of the test phase.

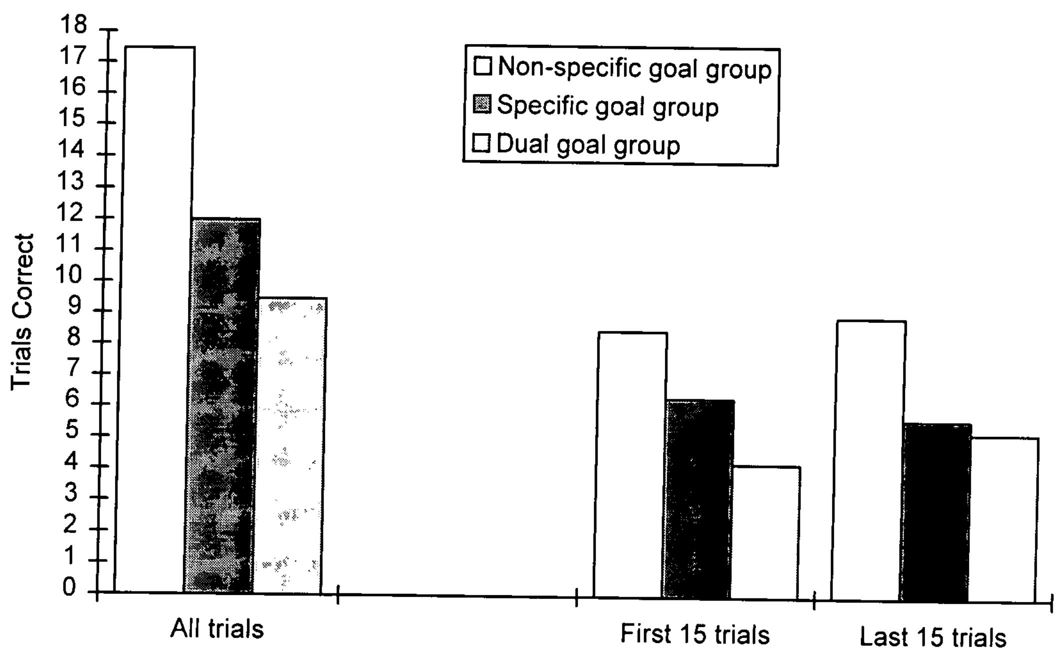


Figure 1.2: Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.

The data in Figure 1.2 were analysed using a 3 (learning goal) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(2,66) = 19.43$, $p < 0.001$. There was no main effect of trial block, $F<1$, and no significant interactions. To confirm how the groups compared to each other during the test phase, planned between groups comparisons were done on the total test phase score and also the scores for the first and last fifteen trials. The non-specific goal group outperformed the specific goal group on all trials, $F(1,44) = 14.27$, $p<0.001$, the first fifteen trials: $F(1,44) = 6.39$, $p<0.02$; and last fifteen trials: $F(1,44) = 13.34$, $p<0.001$. The non-specific goal group also outperformed the dual goal group on all trials: $F(1,44) = 39.17$, $p<0.001$; the first fifteen trials: $F(1,44) = 29.22$, $p<0.001$; and last fifteen trials: $F(1,44) = 18.36$, $p<0.001$. Finally the specific goal group outperformed the dual goal group on all trials : $F(1,44) = 4.23$, $p<0.05$; first fifteen trials: $F(1,44) = 7.96$, $p<0.008$; but not on the last fifteen trials: $F(1,44) = <1$.

Transfer: Another important issue is how the specific and dual goal groups coped when the subjects switched goals from the learning phase to the test phase. Two comparisons for each group were made to examine this issue: (i) A comparison between the number of correct trials during the last half of the learning phase and the first half of the test phase and, (ii) A comparison between the total number of correct trials during the learning and test phases. For comparison (i) a 2 (learning goal: specific vs dual) by 2 (trial block: last 15 trials of learning phase vs first 15 trials of test phase) analysis of variance with repeated measures on the last factor revealed a main effect of learning goal, $F(1,44) = 21.73$, $p < 0.001$: the specific goal group performed better in the last half of the learning phase and first half of the test phase than the dual goal group. There were no other main effects or significant interactions. Therefore, irrespective of learning goal, subjects' performance was not getting better or worse just after their specific goal changed. For comparison (ii) a 2 (learning goal) by 2 (phase score: learning vs test) mixed analysis of variance again showed a main effect of learning goal, $F(1,44) = 20.22$, $p < 0.001$: the specific goal group performed better in the learning phase and test phase than the dual goal group. The interaction between phase score and learning goal just failed to reach significance, $F(1,44) = 2.9$, $p = 0.096$: there was a tendency for the dual goal subjects to improve between learning and test phases. There were no other main effects or significant interactions.

The Prediction Questions

Answers were scored as correct if the response predicted by the subjects was one above, the same as, or one below the response expected from Clegg. The response expected from Clegg was calculated by using the equation from the trial section of the experiment, but not including the random element of that equation. This was omitted as the scoring process took it into account. As mentioned above, if subjects did not get enough questions correct or incorrect during the test trials, it would not have been possible to generate exactly 5 questions of each question type. However, this situation did not arise. On the other hand, there is a potential hazard with the prediction questions that could undermine any conclusions based on their related statistics. Old-correct and Old-wrong prediction questions were selected from the test trials, but any given prediction situation could have occurred more than once and the response on another occasion might have been different from the one given in the selected situation. Therefore, a trial that was selected as an Old-wrong one might have been responded to correctly on another occasion, or vice versa. Such responses (8% of the data) therefore were discarded. The resulting mean percentage of correct responses are shown in Figure 1.3. Due to some of the data being discarded, these results are shown as percentages.

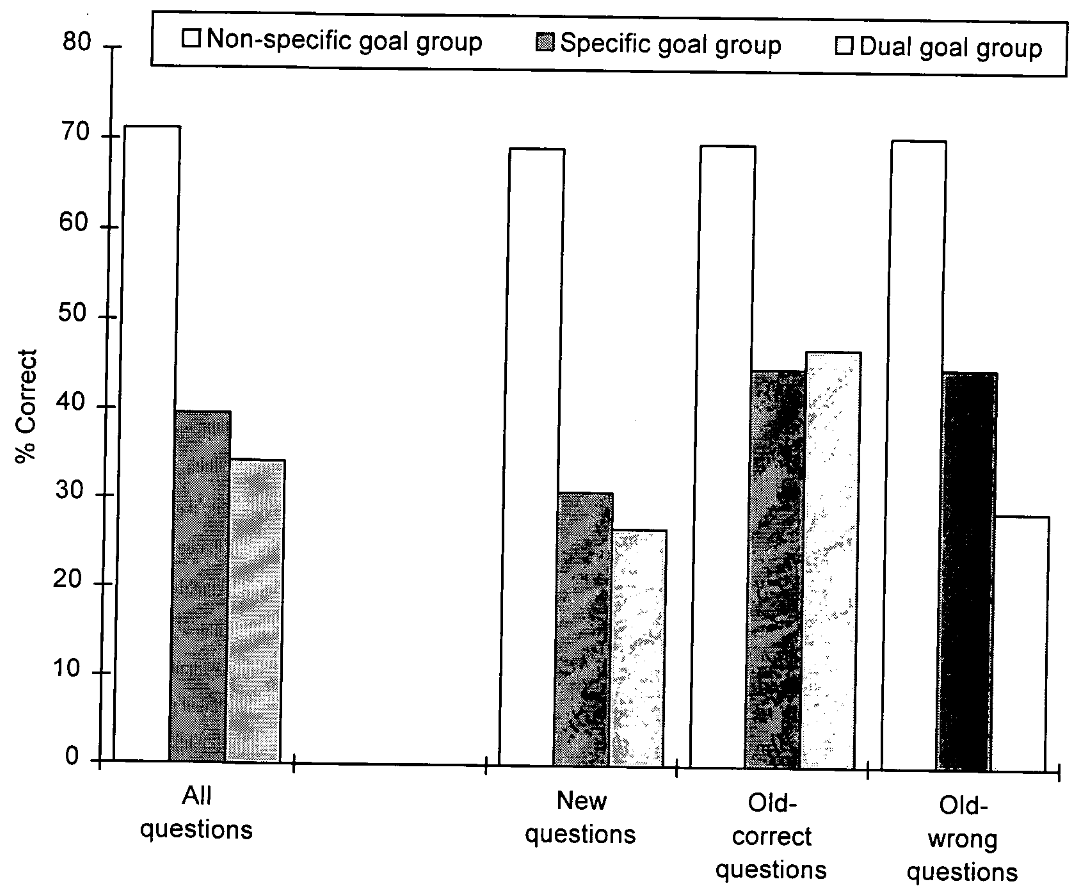


Figure 1.3: Mean percentage of correct responses to each category of prediction questions for each group.

The data in Figure 1.3 were analysed by a 3 (learning goal) by 3 (question type: New vs. Old-correct vs Old-wrong) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(2,66) = 18.05$, $p < 0.001$, a main effect of question type, $F(2,132) = 4.54$, $p = 0.012$. However, the interaction between learning goal and question type just failed to reach significance, $F(4,132) = 2.01$, $p = 0.096$. The lack of significant interaction is surprising. Inspection of Figure 1.3 suggests that all 3 goal groups respond differently to the three kinds of prediction questions. Because of this individual comparisons were carried out on each goal group. Subjects could only score one of six possible values (0-5) for the three prediction question categories. Therefore non-parametric statistics were used when dealing with these results.

To examine how performance varied between the different question types, three comparisons were made separately for each group: (1) A comparison between the percentage of New and percentage of Old-wrong questions answered correctly, (2) A comparison between the percentage of New and percentage of Old-correct questions answered correctly, and (3) A comparison between the percentage of Old-wrong and percentage of Old-correct questions answered correctly.

Wilcoxon matched paired tests showed no significant differences on any of these comparisons for the non-specific goal group. For the specific goal group, New scores were significantly lower than either the Old-wrong scores, $Z = -2.14$, $p < 0.04$, or the Old-correct scores, $Z = -2.32$, $p < 0.03$, but Old-wrong and Old-correct scores did not differ from each other. For the dual goal group, the Old-correct scores were significantly higher than either the Old-wrong scores, $Z = -2.31$, $p < 0.03$, and the New scores, $Z = -2.27$, $p < 0.03$, which did not differ from each other. Thus, the non-specific goal subjects were equally good at answering any of the questions, whether they were Old or New and whether they were Old-correct or Old-wrong. The specific goal group subjects were equally good at answering the Old-wrong and Old-correct questions and better at answering both of these questions than the New questions. And the dual goal group answered the Old-correct questions more accurately than the other two question types, which were answered equally poorly.

To explore how the groups compared to each other, analyses of variance on the total scores showed significant differences between the specific and the non-specific goal groups $F(1,44) = 19.5$, $p < 0.001$ and between the non-specific goal group and the dual goal group $F(1,44) = 26.24$, $p < 0.001$. There was no significant difference between the specific and dual goal groups ($p > 0.2$). The non-specific goal subjects had a higher overall score on the questionnaire than either the specific or the dual goal subjects.

To examine which questions the non-specific goal group were excelling at and whether the specific goal and dual goal group were comparable on all types of questions, three separate between group comparisons were made between each pair of groups; one for each question type. For the comparisons between the non-specific goal and the specific goal groups, there were significant differences on all question types with the non-specific goal group outperforming the specific goal group on each type; for the New scores, U=125, $n1$=$n2$=24, $p < 0.001$, the Old-correct scores, U=154.5, $n1$=$n2$=24, $p < 0.006$, and the Old-wrong scores, U=143.5, $n1$=$n2$=24, $p < 0.003$. For the comparisons between the non-specific goal and the dual goal groups again there were significant differences on all question types with the non-specific goal group outperforming the dual goal group on each type; for the New scores, U=105, $n1$=$n2$=24, $p < 0.001$, the Old-correct scores, U=170.5, $n1$=$n2$=24, $p < 0.02$, and the Old-wrong scores, U=106.5, $n1$=$n2$=24, $p < 0.001$. For the comparisons between the specific goal and dual goal groups, the specific goal group outperformed the dual goal group on the Old-wrong scores, U=173, $n1$=$n2$=24, $p < 0.02$, but performed comparably for the New and Old-correct scores. So, the non-specific goal group was significantly better than both of the other groups on all question types, and indeed, overall on the prediction questionnaire. The specific goal group was significantly better than the dual goal group on the Old-wrong question type.

Another possible problem with the results involving the prediction questions arises from their method of creation. In keeping with the method used by other similar studies (e.g. Berry & Broadbent, 1984), the prediction questions were defined from a trial situation that used the subject's previous input, Clegg's previous output, and the subject's new input. There is no sure way of knowing how subjects may have been defining situations themselves during the trials.

One plausible alternative is that subjects defined a situation by using only Clegg's previous output and the subject's new input. This is all the information that is needed to make an accurate prediction of Clegg's new output. With this possibility in mind the results were reanalysed defining the prediction question situations by the last two elements (again discarding questions for which the situations could have occurred with both a correct and an incorrect response). The results can be seen in Appendix 1, pg. 230. These reanalyses showed an identical pattern of statistics with all the results that were significant before remaining significant (all p values < 0.05).

Correlations between control performance and predictions: To examine how much the questionnaire scores relied upon control performance, total prediction questionnaire scores were correlated with the number of correct trials during the test phase. Spearman rank correlation coefficients were 0.26 for the specific goal group, 0.87 for the non-specific goal group and 0.38 for the dual goal group. These correlations were significant for the non-specific (p < 0.001) but not for the specific goal or dual goal groups. This suggests that for the non-specific goal group, predictions were directly related to control performance in the test trials.

The questionnaire scores for the specific and dual goal groups were also correlated with the number of trials correct during the *learning* phase. The correlation coefficient was -0.14 for the specific goal group and -0.05 for the dual goal group. Neither correlation was significant.

The Rule Description Questions

Subjects' answers to the two rule description questions (asking how to control Clegg and asking what was Clegg's underlying pattern) were treated together as subjects generally answered only one of the questions and included information in that answer that was relevant to both questions. The answers were judged by two judges and placed into one of three categories; *No information or Wrong, Partially Correct, Correct.* Answers were categorised as *No information or Wrong* if subjects gave no relevant information as to the pattern Clegg was following or as to how they controlled Clegg. Answers were also assigned to this category if part of the answer gave wrong information. Answers were categorised as *Partially Correct* if subjects: mentioned Clegg's tendency to move along the scale beyond the subject's response (away from his own); mentioned any other information that described this approximate characteristic of Clegg's behaviour; made one precise possible prediction of Clegg's behaviour; or mentioned how Clegg's behaviour clustered around a continuous behaviour of the subjects. Answers were categorised as *Correct* when subjects mentioned Clegg's tendency to move along the scale, beyond the subject's response (away from his own) AND described the distance along the scale that Clegg would move (i.e. roughly double the distance the subject was from Clegg). Answers that made 3 or more precise possible predictions of Clegg's behaviour were also ranked as *Correct*. Examples of the different responses are shown in Appendix 7, pg. 261. Both judges categorised the answers identically. These categorisations can be seen in Table 1.2.

Table 1.2
Numbers of subjects giving each category of response to the rule description questions

Category	No information or Wrong	Partially Correct	Correct
Group			
specific goal	**20**	**3**	**1**
non-specific goal	**5**	**4**	**15**
dual goal	**22**	**2**	**0**

As can be seen from the data in Table 1.2, the specific and dual goal groups achieved far fewer answers in the *Correct* category and far more answers in the *No information or wrong* category than the non-specific goal group. Fisher exact probability tests comparing the number of answers in the *No information or wrong* category and in the *Correct* category showed these differences to be highly significant; with the specific goal group $p < 0.001$, and with the dual goal group $p < 0.001$. There was no significant difference between the specific and dual goal groups. Thus, the non-specific goal group was better than the other two groups at producing answers containing declarative knowledge. It might be argued however, that too strict a criterion was used to categorise *Correct* responses and that with a looser criterion the non-specific goal group might have been more similar to the other two groups. With this possibility in mind the Fisher exact probability tests were repeated, but this time the number of questions in the *Partially Correct* and *Correct* categories were added together. Once again the tests showed that there were highly significant differences between the non-specific goal group and the other two groups ($p < 0.001$).

DISCUSSION

The aim of the present experiment was to test the proposition that a specific goal leads to instance learning of a dynamic control task, while a non specific goal leads to rule learning. The results show that goal specificity does indeed determine whether instance learning or rule learning occurs. When subjects are given a specific goal, instance learning is observed, consistent with previous studies of implicit learning (e.g. Berry and Broadbent, 1984, 1988). But when subjects are given a non specific goal, rule learning is observed, consistent with studies of explicit learning (e.g. Owen and Sweller, 1985; Sweller, 1988; Vollmeyer et al, 1996).

The conclusion that the specific goal group learned instances is deduced from the following points: (1) the significant difference in performance on the different prediction questions, namely, poor performance on the New questions; (2) the poor answers to the rule description questions; (3) the lack of a significant positive correlation between total questionnaire score and trial performance. The conclusion that the non-specific goal group learnt rules is deduced from the following points: (1) the similarity of performance on all prediction question types, particularly, being able to do just as well on New questions as on the Old-wrong and Old-correct questions; (2) the success at answering the rule description questions; (3) the significant positive correlation between total questionnaire score and test performance. The notion of the two groups learning differently is further reinforced by the comparisons between the two groups. The non-specific goal group performed better on the rule description questions, indicating the use of explicit rule based knowledge.

These results agree with those showing effects of goal specificity in studies of explicit learning (e.g. Sweller, 1988; Vollmeyer et al, 1996); they extend those results by showing comparable findings in an implicit learning task. Whittlesea and Dorken (1993) found that what is learned about instances depends on the learning goal or the purpose for which the task was performed. Those results are extended too by showing that the learning goal can also affect whether the subject learns instances or rules.

It has been proposed that the non-specific goal group learnt explicit rules. However, an alternative explanation is that the group learnt by constructing a look-up table in the same way as the specific goal group but the lack of a specific goal led the look-up table to having a more varied set of entries. The situations in the prediction questions would then have a higher chance of being the same as, or similar to, the entries in their look-up table than if the table had been created while pursuing a specific goal. Such an explanation is compatible with the results of Buchner et al (1995) where there was a significant positive correlation between prediction success and the number of state transitions experienced during learning. This alternative explanation might explain the non-specific goal group's better performance on both new and old prediction questions. However, the suggestion of learning by the use of a look-up table implies that the non-specific goal group was not using rule constructing strategies. Two arguments are contrary to this suggestion; First, the significant positive correlation between total prediction scores and trial performance has not been found with person interaction task studies claiming implicit learning. Second, it is clear from the answers to the rule description questions that the non-specific goal group had a good verbalisable understanding of Clegg's pattern; they knew the rule he was following. Thus, the evidence is in favour of the proposal that the non-specific goal group was learning explicit rules.

There are two possibilities as to why having a specific goal prevents rule learning while a non specific goal encourages rule learning. One is that the more varied set of trial situations that the non-specific goal group is likely to encounter yields rules about state transitions. This more varied set of trials may make the underlying pattern more salient. The second possibility is that the non-specificity of the learning goal encourages explicit hypothesis testing. This second possibility is consistent with the dual space model. Vollmeyer et al suggest that the lack of a specific goal means that subjects have no guidance as to how to search the instance space. Therefore subjects may use exploration of rule space to direct their search in instance space. Such a view does not depend on subjects encountering a more varied set of trial situations for rule learning to occur but assumes that the non-specific goal results in a search of the hypothesis space. It is important to establish which of these reasons may be responsible for inducing explicit rule learning as the latter suggestion implies that specific goals may encourage instance learning and in some situations, where transfer is needed for example, this may not be desirable. The experiment reported in chapter 2 suggests that the reason subjects learn rules is because they have a non specific goal and not because they encounter more varied information as a result of meeting the non-specific goal.

A dual goal group was also included in this study. When the results of this group are compared to those of the specific goal group, it was found, as expected, that the specific goal group was better at both learning and test than the dual goal group, and that both groups were equally poor at giving an explicit description of the underlying rule, indicating an absence of explicit rule learning whenever there is a specific goal. However, the results also showed that while the specific goal group made better predictions in both old correct and old wrong situations compared to new situations, the dual goal group only did well in old correct situations, being equally poor in old wrong and new situations.

This ability of the dual goal group subjects to predict correctly in old correct situations indicates that some consistent instance learning has occurred. However, the inability of these subjects to predict correctly in old wrong and new situations indicates that rule learning has not occurred. Evidently, working memory capacity is overloaded by having two goals, and so rule learning is prevented. However, despite the high cognitive load, instance learning seems unimpaired. This latter claim assumes that instance learning leads to the development of a look-up table in which correct trials are stored, since dual goal subjects correctly predicted Clegg's response in old correct situations only. Such an assumption also implies that the specific goal group subjects learned to control Clegg using a combination of instance learning and rule learning, their ability to predict correctly in old correct situations being due to instance learning and their ability to predict correctly in old wrong situations being due to rule learning. This possibility is consistent with the findings of Dienes and Fahey (1995). They tested a look-up table model based on Logan's (1988) instance theory and found that the best fit with the data occurred when the look-up table stored correct responses only. On the other hand, a model of explicit learning, based on strategies suggested by an independent group of subjects, gave the best fit to the data on predictions in old wrong situations.

Such a view also seems plausible from a processing point of view. The specific goal is to make the computer produce and maintain a specific response. If a subject is simply trying to achieve this goal, it makes sense to store only correct trials because these are the trials that indicate how to reach the goal. This is what happened in the dual goal group. By contrast, if subjects spontaneously generate and test hypotheses about the best way to reach the goal, then it makes sense to store incorrect trials as well, because incorrect trials indicate failed hypotheses. This is what happened in the specific goal group. The above view also implies that correct instances are learned implicitly, since it argues that learning of correct instances was not affected by the high cognitive load imposed by having two goals. However, since both goals were task related, this proposal needs to be explicitly tested by asking subjects to carry out a neutral concurrent task at the same time as fulfilling their assigned goal or goals. Such a test would establish more clearly whether learning correct instances really is implicit, since implicit learning, but not explicit learning, should be preserved under a neutral concurrent task. This idea is explored in Experiment 4.

The finding that initial learning was better in the specific goal group than the dual goal group is consistent with the results of Berry and Broadbent (1988). They gave subjects a specific goal for 20 trials and then for the next 20 trials they told subjects that they should also try to find the pattern underlying the computer's behaviour. This additional non-specific goal caused a severe decrement in their learning. Vollmeyer et al's (1996) specific goal subjects could also be regarded as a dual goal group. All their subjects were told to explore the system to learn as much as possible. The specific goal subjects were then told in addition what specific goal they would be tested on in the test trials. So, in effect, the specific goal subjects learned under both a non specific goal and a specific goal. In that study, the specific goal group and the non specific goal group learned to reach the specific goal equally well; a result that is contrary to these results and those of Berry and Broadbent. The most likely reason for the discrepancy with the results shown here is that Vollmeyer et al's subjects did not have to make the computer give the specific response until the test trials and so exploration of rule space during learning would have been less constrained than in the present study. In Berry and Broadbent's study, the initial learning was under a specific goal and it is this initial learning that was disrupted by the later attempts to search for the underlying pattern.

One thing that is clear in the results of both the specific goal and the dual goal groups is that these subjects do not learn the rule underlying the computer's behaviour. This finding is in accord with Sweller's (1988) and Owen and Sweller's (1985) observations on trigonometry learning, where learners who had a specific goal did not transfer their learning to novel situations while learners who had a non specific goal did. Sweller argues that problem solving via means-ends analysis prevents rule learning because of the cognitive load imposed by having to meet a specific goal. According to Sweller, poor transfer with specific goals arises because having to monitor goals increases the cognitive load of the task, and so reduces the capacity available for rule induction. However, the results presented here suggest that pure instance learning only occurs when the subject has to meet two inconsistent goals. When subjects only have to a meet a specific goal, a mixture of instance learning and rule learning occurs, as observed also by Buchner et al (1995) and Dienes and Fahey (1995).

According to the dual space model, a specific goal encourages the use of means-ends analysis for finding a route through instance space, thus precluding a search of rule space and full rule learning. When explicit rule learning occurs with a specific goal, a minimal search of rule space allows the subject to generate and test heuristic strategies in explicit attempts to reach the goal. These strategies tend to be simple rules of thumb that focus on how to reach the goal; for example, to keep responding in the *Polite* region (Dienes and Fahey, 1995). The strategies are not hypotheses that explain the general pattern without concern for the goal. Thus it is probable that the explicit strategies used by specific goal subjects during instance learning are less than optimal because the need to reach a specific goal always brings the subject back to instance space. Free exploration of rule space unimpeded by the need to find a route to a specific goal is necessary if the correct rule is to be found.

On this account, therefore, it is not cognitive load that accounts for poor rule learning with specific goals, but the minimal use of rule space due to the need to concentrate on instance space in order to find a route to the goal. The lack of any rule learning when there are two goals may be attributed to the cognitive load imposed by simultaneously having to search rule space for a suitable strategy to reach the goal and to search it to find a suitable hypothesis about the underlying pattern of the task. In addition, the former search always takes the learner back to instance space because the goal has to be reached, and so precludes the free exploration needed to find an underlying pattern.

In addition to the theoretical issues discussed above, these results also have methodological and practical implications. Regarding methodological implications, it was noted in the introduction of this experiment that it has proven very difficult to demonstrate the existence of implicit learning. Shanks and St. John argue that only one study (Hayes and Broadbent, 1988) has come close to showing a double dissociation between experimental variables and implicit vs. explicit learning. Hayes and Broadbent argued that a person-computer interaction task that had a salient rule was learned explicitly while a task with a non-salient rule (comparable to the task used in this study) was learned implicitly. They then showed that when the equation was changed without warning, the performance of subjects given the non-salient rule fell significantly below that of subjects given the salient rule. In a subsequent experiment, subjects completed the same task while also carrying out a concurrent task. In this experiment, relearning after the equation change was worse for subjects with the salient rule than for those with the non-salient rule. Thus, they appeared to have shown a double dissociation: only implicit learning was disrupted by the simple equation change, while only explicit learning was disrupted by a concurrent task.

However, Green and Shanks (1993) failed to replicate these results and suggested, on the basis of additional results, that the failure to replicate could be explained by saying that the non-salient task was more difficult than the salient task and that this difference, not differences in the way the two tasks were learned, was responsible for all the results. When two different tasks are used for implicit and explicit learning, it is always possible that task differences might account for the results rather than differences in learning processes. In the experiment reported here, different learning processes were attributed to the same task as a function of the learning goals. The differences that were observed in test performance, prediction questions and explicit questions, therefore, can be attributed to these different learning processes, either implicit, explicit, or a mixture of the two. Using manipulations that might have differential effects on test performance and on the questionnaires may be a better way of investigating the two learning systems and the relationship between them than trying to show a dissociation between them.

As regards practical implications, the results show that not only does rule learning result in better understanding of the system than instance learning as would be expected, it also results in better control of the system, that is, better problem solving. The results, therefore, suggest that an important method for encouraging learning in the classroom is to give learners tasks that have non-specific goals so that rule learning can be facilitated. As an example, Schoenfield (1985) has pointed out that much of mathematics learning in schools is based on practise at solving problems. He comments that such an emphasis fails to encourage transfer to novel situations. This study suggests one way in which problem based learning that does facilitate transfer might be achieved.

In conclusion, the results indicate that goal specificity determines whether rule learning or instance learning occurs. Subjects given a specific goal learn instances while subjects given a non specific goal learn rules. Subjects given both goals also appear to learn instances, but only memorise correct trials. On the basis of these results, it has been speculated that subjects given a specific goal learn by using a mixture of implicit and explicit processes while subjects given a dual goal only learn implicitly. However, more work is needed to test this possibility, since it assumes that dual goal subjects do not engage in any explicit learning at all (for this see experiment 4, pg. 103). The results were also discussed in relation to dual space models in which problem solving and hypothesis testing are explained within a single theoretical framework. It was suggested that means-ends analysis, induced by a specific goal, takes place in instance space with minimal use of rule space while in rule learning, search of rule space predominates. Finally, comments were made on the benefits of using a manipulation that affects the type of learning subjects engage in but does not alter the structure or difficulty of the task they have to learn.

Chapter II

Experiment 2: Learning Goals - A Direct Influence On Cognitive Processes Or Just On Salience?

Chapter 1 reviewed some of the work on learning systems. The implicit explicit learning distinction was discussed. An area of common agreement regarding the learning systems was that, ignoring the issue of consciousness, implicit learning can be described by instance models of learning whereas explicit learning can be described by hypothesis testing and rule induction models of learning. The data from Experiment 1 also suggested that there may be some examples of learning that exhibit behaviour that is a combination of implicit *and* explicit instance learning. It was argued that the dual space model of learning provides a theoretical framework that can incorporate and enhance both instance learning and rule learning. The literature reviewed suggested that the forms of learning are differentially sensitive to the salience of the material that is learnt (e.g. Berry & Broadbent, 1988), with instance learning occurring for low salience material and rule learning for high salience material. However, the results of Experiment 1 demonstrated that a more fundamental influence upon learning mode is related to the learning goal given in the initial instructions to subjects. If subjects are given non-control oriented, pattern search instructions, then in the case of the person interaction task they can learn in an explicit manner even though the pattern is non-salient. The aim of the study described in this chapter is to establish why, despite the pattern being non-salient, pattern search instructions lead to explicit learning. One distinct possibility is that pattern search instructions simply increase the salience of the material to be learnt by guiding subjects to sample a wider range of information about the person interaction task. This would suggest that the fundamental instigator of mode of learning is still related to the salience of the material to be learnt. The dual space model of learning described in chapter 1 would suggest otherwise. In line with this model it is proposed that explicit learning arises due to some more intrinsic characteristic of the instructions - namely the lack of emphasis on control, and the subsequent alternative cognitive processes that this allows. In terms of the dual space model, the lack of a specific goal allows subjects to spend more time exploring rule space.

The introduction is organised as follows. First, the key findings from Experiment 1 are reviewed; namely, that the fundamental reason why the person interaction task is learnt in a way best described by instance models of learning is related to the subjects' control task goal that they are given during their learning phase; subjects were asked to control the system to make it produce a specific output. If, however, instructions were changed to make subjects deliberately search for an underlying pattern to the system, without the hindrance of a control goal, then subjects were much more likely to explicitly learn the underlying rules that govern the system. Next, two possible reasons as to why explicit rule learning is induced are presented and a paradigm proposed (the observation paradigm as reported by Berry, 1991) for testing the two possibilities. Finally, the experiment that will examine exactly why pattern search instructions lead to explicit learning is introduced. The experiment also attempts to replicate Experiment 1's original findings under stricter, more constrained conditions.

In all previous experiments using the person interaction task, the subjects have been instructed to control the computer person so that it produced a specific output. In chapter 1 it was suggested that this control goal might be the critical factor in leading subjects to learn the task implicitly. This suggestion was tested by comparing one group of subjects given the normal instructions (the control task group) with another group of subjects given no control goal (the pattern search group) but simply told to establish the pattern that the computer person was following. These groups differed by only one sentence in their initial instructions. Subjects performed 30 learning trials then another 30 test trials where their control performance was measured. Finally their verbalisable knowledge was examined via a questionnaire consisting of rule description questions and carefully constructed prediction questions. The results clearly demonstrated that the pattern search subjects learnt explicitly accessible rules while the control task subjects learnt in a way that could be described as instance learning. These conclusions were deduced from the fact that the pattern search subjects outperformed the control task subjects on the prediction questions, particularly on the prediction questions where subjects made predictions from situations they had never encountered before. The pattern search subjects also outperformed the control task subjects on the rule description questions with at least half the subjects explicitly describing the pattern they had learnt (compared with none in the other group). Finally unlike the control task subjects, the pattern search subjects showed a positive correlation between prediction question performance and control performance. Hence, from this study it can be concluded that something critical to the control goal given to subjects in the control task group leads them to learn not by explicit hypothesis testing and rule deduction but by memorising instances.

One way to understand the above conclusion is to look at the other side of the coin and examine exactly why the pattern search instructions lead to explicit rule learning. It is known that lack of salience of a pattern encourages instance learning. One obvious possibility therefore (the 'salience explanation'), is that subjects with no control goal view a wider range of interactions than subjects with the control goal and this wider range of interactions simply increases the salience of the underlying pattern and hence induces explicit learning. If this is the reason then it reinforces salience as the critical factor in inducing one type of learning over another.

This notion of a wider range of interactions being responsible for differing behaviour has support from a recent study by Buchner, Funke, & Berry (1995). Their study demonstrated that subjects who successfully attained the specific goal did poorer on prediction questions following the task. Their explanation was that these subjects achieve the specific goal more often than less successful learners and therefore saw less of the system thus reducing their ability at the prediction questions. However those that were poorer at achieving the specific goal were better at answering the questionnaire because they had seen more of the system. Buchner et al, specifically examined the amount of state transitions that subjects encountered, and found a negative correlation between state transitions encountered and goal performance, and a positive correlation between number of state transitions encountered and prediction question performance.

In the dynamic system that Buchner et al's subjects learnt there were a finite number of state transitions. A state transition is a specific combination of input and output. In exploring how wide a range of interactions subjects encounter it is possible to measure the number of state transitions that they encounter and also the number of times a particular state is encountered. In the case of the person interaction task a state is simply a behaviour, and a state transition is a specific combination of output behaviour from the computer person and an input behaviour from the subject. In Experiment 1 it is likely that the group with the non specific goal encountered both a larger number of states and also a larger number of state transitions, potentially increasing the salience of the pattern they were learning.

Another reason as to why the pattern search instructions lead to explicit learning may be related to the fact that these instructions lack any emphasis on specific control. The lack of emphasis on achieving a specific goal should guide subjects away from instance space to explore rule space. In other words the salience of the material is not relevant, it is the cognitive processes that are triggered by the non-specific pattern search goal that produce rule learning. Applying this line of thought to the control task group suggests that the presence of a specific control goal is a critical factor in inducing instance learning. This shall be referred to as the 'goal explanation'. If the 'goal explanation' is correct then the nature of the cognitive processes induced by a specific control goal on the one hand and a non-specific, pattern search goal on the other, would merit further study.

An ideal paradigm for distinguishing between the 'salience explanation' and the 'goal explanation' is the observation paradigm as used by Berry (1991). Ideally subjects should be given the pattern search instructions and also somehow have their range of interactions with the computer person limited to that of a subject given control specific instructions. If this group of subjects did not learn explicitly then the 'salience explanation' could be used to explain the rule learning of groups with the pattern search instructions. Alternatively, if they were successful at learning explicit rules then the 'goal explanation' could be considered. The observation paradigm has subjects' initial learning experience come from purely observing some earlier interactions. Subjects learning from this observing experience can then be tested by examining both their control performance and verbalisable knowledge in the same ways as was done in Experiment 1.

To test between the two explanations a group of subjects can simply be given pattern search instructions and make the learning phase (the first 30 trials) consist of the subjects observing the learning phase of a group that was learning the person interaction task under specific goal instructions. Additional to this experimental group a number of control groups are needed. Aside from having a complete design, these control groups allow alternative explanations of the results to be ruled out. Some consideration to these alternative explanations is needed.

When Berry used the observation paradigm she found subjects' learning to be very poor following their observing. However, unlike the experimental group proposed here, Berry gave her observers specific goal instructions. Still, in light of the apparent difficulty that observing causes learners it would be sensible to compare the experimental group with a control group that were observing a group of subjects that had also been given pattern search instructions. This way, the potentially detrimental effect of observing on learning would be controlled for.

If it was found that the two observing groups were learning explicit rules as predicted, it would be possible to argue that the very act of observing leads to explicit learning. This would prevent the conclusion that rule learning is directly produced from having pattern search instructions. To rule out the notion that the act of observing stimulates rule learning, situations need to be shown in which observing doesn't lead to rule learning. It is unlikely that the act of observing is causing rule learning as Berry (1991) did not find this to be the case. However, there are slight differences between the procedure used in this experiment and Berry's procedure, so it is sensible to test for this possibility. Therefore in the experiment reported below two additional observing groups of subjects are included who are given non-pattern search, control task instructions, thus allowing one to see whether the pure act of observation leads to explicit learning.

This experiment shall also attempt to replicate the key elements of the original findings (as reported in Experiment 1 and in Geddes & Stevenson, in press) under stricter conditions. With this attempt successful, the data can then be used for the 'observing groups' to observe. Also, the range of interactions that the two replication groups view can be examined and it can be seen if they are different and therefore if the 'salience explanation' is feasible. In Experiment 1 the control task models and the pattern search models performed 30 trials of a learning phase where their goals were different; control and pattern search instructions respectively. The groups then had 30 trials of a test phase where their goals were identical; a new, specific control goal. The questionnaire then followed this test phase. It was from this test of verbalisable knowledge that the critical claim of the pattern search group learning explicitly accessible rules was able to be made. A stricter condition would be to give the questionnaire immediately after the learning phase thus not allowing it to measure any of the learning that might have occurred in the 30 trials of the test phase. There is another reason to redesign the experiment in this way: Observers are supposed to be learning from their observations during the learning phase. It will be vital to have their verbalisable knowledge tested immediately after their observation period and not following the subsequent test of their control performance. Otherwise, any success on the questionnaire may simply reflect what they had learnt during the test trials not what they learnt during their observations. Thus the questionnaire was placed immediately after the learning phase for both the models and the observers.

Hypotheses: For the two groups that will be observed (the models) the following results are predicted. The pattern search models should learn explicitly accessible rules and hence they should perform equally well on all prediction situations (including the new situations). There should be a significant positive correlation between control performance and total questionnaire score. There should also be evidence of verbalisable knowledge in the answers to the questions about the underlying rule.

The control task models should learn by memorising instances and should meet the expectation of learning by building a look-up table. Therefore subjects should perform worst on new situation prediction questions. There should be no significant correlation between control performance and prediction performance.

In comparisons between these two groups, the pattern search model group should perform better overall on the prediction questions and, specifically, better on new prediction situations. The pattern search models should give better rule descriptions.

Additionally the range of interactions encountered by these two groups should differ. The control task models should have a narrower range of interactions than the pattern search models. With the control task models having a specific goal, a larger part of the interactions should revolve around the goal. More specifically, the control task models should encounter fewer state transitions than the pattern search models and the control task models should encounter a less even distribution of states.

The observers given pattern search instructions during their observation period should learn accessible rules irrespective of the goal of the model they are observing. Therefore their test results should be the same as those predicted for the models who have pattern search instructions.

The observers given the control task instructions during their observation period should differ considerably from both the observers and models given pattern search instructions. The results may match those of the models given control task instructions implying that these subjects learnt instances or, alternatively, the results may indicate that no learning took place. To test for this latter prediction, key results will be compared to those expected by chance.

METHOD

Subjects: The 72 volunteer subjects were Durham University graduate and undergraduate students, aged between 18 and 24. As for the subjects in every experiment in this thesis, this was the only experiment that they had ever done that featured a dynamic system control task.

Design: A 2 (goal of model) by 2 (goal of observer) independent groups design was used. Both models and observers were given either a control task goal or a pattern search goal. Half of the control task observers observed control task models and half observed pattern search models. Similarly, half of the pattern search observers observed control task models and half observed pattern search models (see Table 2.1).

Table 2.1 - Labeling of Observer groups

	Goal of Models			
	Control task		Pattern search	
Observing group name	**CTCT**	**CTPS**	**PSCT**	**PSPS**
Goal of Observers	Control task	Pattern search	Control task	Pattern search

The first 2 letters of the name identify the goal of the model, the second 2 identify the goal of the observer. CT = Control Task; PS = Pattern Search

The first 24 subjects were randomly allocated to one of the two groups of models - the control task models or pattern search models. This had to be done first as the remaining subjects needed to observe the models' data. The remaining 48 Subjects were then randomly allocated to one of two observing groups - the control task observers or the pattern search observers. Within each of these groups subjects were randomly allocated to either a group that observed control task models or a group that observed pattern search models, as shown in Table 2.1. Each subject in the observing groups was then randomly assigned to a model with the relevant goal. It would be this model's data that the observer would encounter during their learning phase of the experiment. Models interacted with the computer while learning; observers observed the learning responses of the models. For reference in this chapter the four sets of observers are referred to as; CTPS observers for the pattern search observers observing the control task models; PSPS observers for the pattern search observers observing the pattern search models; CTCT observers for the control task observers observing the control task models and PSCT observers for the control task observers observing the pattern search models.

All subjects were required to complete 30 trials for the learning phase. All subjects were then given the unexpected questionnaire. This consisted of 20 prediction questions followed by a rule description question section which consisted of 2 questions for the control task models and 1 question for the other groups. There were 3 different types of prediction question for the control task models - 5 based on old correct trials, 5 on old incorrect trials and 10 on new trials. For the other groups (who had no initial task that they could get correct or incorrect) there were only two types of prediction questions - 10 based on old trials and 10 based on new trials. For all groups the 20 prediction questions were displayed in a random order. Following the questionnaire, subjects' control ability was then examined during the test phase of the experiment. This consisted of a final 30 trials.

The task : Subjects learned the Clegg version of the person interaction task which was identical to that described in chapter 1. Instead of interacting with the computer, observers were told to press the space-bar to view the model's next input and Clegg's output. This allowed the observer to view an identical amount of information on the screen as the model was able to view (6 interactions at any one time).

For the learning phase the subjects were given virtually identical instructions. The identical part of the instructions concerned the information given in the previous paragraph. Additional to this, the control task models were told "Your aim is to shift Clegg to the '*Polite*' level and maintain him at that level". The pattern search models were told "Your aim is to establish under what pattern Clegg is reacting". The pattern search observers were told that, "Your aim is to establish under what pattern Clegg is reacting. However, for this section of the experiment you will not interact with Clegg, but will view some interactions that have occurred". The control task observers, in addition to having the general task explained to them, were told that, "For this section of the experiment however, you will not interact with Clegg, but, will view some interactions that have occurred. You should watch what the person has done on the earlier occasion as this should give you a feel for how Clegg responds. It is important you pay close attention to the interactions you shall be viewing as, later, you will have to control Clegg, making him produce a specific output and then maintaining his output at the specific level." The observers were not told the task that the model had been set. To remind subjects of their respective goals, the goal of their task was permanently displayed on a piece of paper attached to the bottom of the screen. The rest of the experiment was identical for all four groups.

For all the groups, the test phase was identical to the learning phase for the control task models except that the goal was different. The new goal for all six groups was to make Clegg be *Very Friendly* and maintain him at that level. Clegg's response of *Friendly, Very Friendly* or *Affectionate* was scored as being on target to allow for the random element in the equation.

The Questionnaire

The prediction questions: There were 20 prediction questions. Each question took the following form: First there was a typical trial situation presented. The subject's and Clegg's behaviour was displayed on the screen, below this the subject's new behaviour was displayed -

e.g. You were *Very Cool,* Clegg then was *Very Rude,* You were then *Polite.* Subjects then had to predict what Clegg's response would be.

The prediction situations were generated in four different ways:

A) New situation; The situation was generated randomly from a list of all possible situations that the subject had not encountered during the learning phase.

B) Old situation; The situation was randomly selected from all the trials during the learning phase.

C) Old-wrong situation; The situation was randomly selected from all the trials which the subject had got wrong during the learning phase.

D) Old-correct situation; The situation was randomly selected from all the trials that the subject had got correct during the learning phase.

The control task models were given ten New situations, five Old-wrong and five Old-correct situations to predict from. To produce all five situations of either Old-wrong or Old-correct, meant that the subject must get at least five trials wrong or five correct respectively during the test phase. The program controlling the experiment allowed for the possibility of this not occurring and would have substituted any uncreated situations with New situations. Ten New and ten Old type situation were given to the other groups to predict from.

The rule description questions: One question was presented and subjects were asked to answer it freely, not worrying about wording or grammaticality. The question was:

"Could you try to describe what sort of pattern you thought Clegg was using to respond to your behaviour?"

The control task models were also asked "How did you get Clegg to behave as you wanted him to?" This extra line of questioning was included for the control task models as the question may be more suited to tapping any explicit knowledge they had acquired as a result of their particular learning goal.

Procedure: First, subjects were randomly allocated to one of the six groups. As mentioned above, the groups received virtually identical initial instructions. The differing part of the instructions dictated the aim for that particular group in the learning phase. Apart from this, the remainder of the experiment was identical for all the groups:

The instructions explaining the nature and aim of the subjects' initial learning task were presented first. These were followed by the learning phase of the experiment. On completion of this phase subjects were presented with instructions for the prediction questions. These instructions simply described the nature of this new task and provided an example of the information from which they would have to make a prediction. They also explained that each question was unrelated to the previous one. After completing the prediction questions subjects were given a pen and paper and were asked to answer the rule description question(s). Following the questionnaire subjects received instructions describing their new aim for the test phase and then the test phase started. For both the learning and test phases Clegg initiated the interactions by displaying one of the three adjectives centred on *Polite*.

Throughout the experiment, all instructions appeared on the computer screen but were also read out to the subjects. The experimenter stayed with the subject throughout the experiment in order to answer any arising questions.

RESULTS

Performance during learning and testing

As mentioned before, trials were scored as correct for the control task models if they got a response from Clegg of *Indifferent, Polite* or *Very Polite* during the learning phase of the experiment. This takes into account the random element of the equation producing Clegg's behaviour. Due to the lack of specific aim for the other groups during their learning phase, no measure could be made for their performance during this first set of trials. For all groups, trials were scored as correct during the test phase if subjects got a response from Clegg of *Friendly, Very Friendly* or *Affectionate.*

The total number of correct trials during each phase of the experiment and the total number of correct trials for each half of an experimental phase were measured. The mean numbers of correct trials were calculated for all groups in each of these categories.

The Models

Initially the results are examined for the control task and pattern search models. These are the results that the four groups of observers shall view. It is important to first establish that the models are learning in the predicted manner so that one is sure that the observers are observing models who have learned in distinctly different ways.

The Learning Trials

The control task models achieved a total score of 9.25, which was significantly above that expected by chance (a value of 7.4), $t(11) = 4.32$, $p < 0.001$. Chance level was calculated by running 50,000 simulated sessions, each of 30 trials, in which the subjects chose any one of the 12 responses with equal probability. All values of chance throughout this thesis are calculated from this method.

A within groups comparison was made for the control task models comparing the 1st fifteen trials (with a score of 3.33) with the 2nd fifteen trials (5.92). A paired sample t-test showed a significant difference between these means; $t(11) = -2.35$, $p < .04$. This shows that the control task models were producing more correct trials towards the end of the learning phase than at the beginning. A comparison was not made for the pattern search models as no measure of their performance was possible.

The Test Trials

The mean number of correct trials for both groups of models, for the entire test phase and each half of the test phase can be seen in Figure 2.1. For both groups, their mean total scores were significantly above that expected by chance (a value of 7.4), both p values < 0.03. The data in Figure 2.1 were analysed using a 2 (learning goal) by 2 (trial block: first 15 trials vs last 15) analysis of variance with repeated measures on the last factor. The results revealed a main effect of trial block, $F(1,22) = 4.62$, $p = 0.043$. There was no main effect of learning goal. The interaction between learning goal and trial block just failed to reach significance, $F(1,22) = 3.74$, $p = 0.066$. (See Appendix 2 for the ANOVA tables and full sets of t-tests for this experiment, pg. 233). It is clear then that overall during the test phase the two different learning goals did not affect performance during the test phase. However, during the test phase there was an overall improvement from the first half to the second. The marginally significant interaction suggests that this improvement was stronger in one of the groups. Inspection of Figure 2.1 indicates that the improvement during the test phase occurred mainly for the control task models.

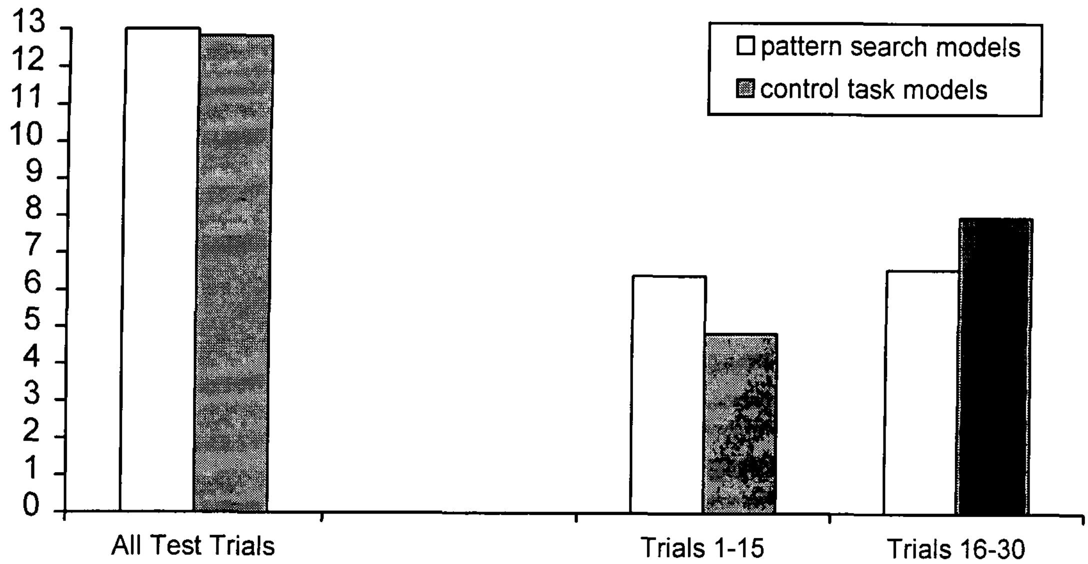


Figure 2.1: Mean number of correct trials in the test phase for the models. Data are shown for all 30 trials combined and for the first and second 15 trials.

Transfer: Another important thing to examine out of the trial performance data is how the control task models coped when they switched goals from the learning phase to the test phase. Two comparisons were made to examine this: (i) A comparison between the number of correct trials during the last *half* of the learning phase and the first *half* of the test phase and, (ii) A comparison between the *total* number of correct trials during the learning and test phases. Paired sample t-tests showed no difference for comparison (i) $t(11) = 0.82$, $p=.428$, but a significant difference between values for comparison (ii), $t(11) = -2.33$, $p < 0.05$. This indicates that there was no initial decrement in performance with the change in goal, but performance improved overall between the learning and test phase.

The Prediction Questions

Answers were scored as correct in the same way as that described in Experiment 1. As in Experiment 1, responses were discarded from the control task models (5% of the data) if a selected trial type (e.g. Old-wrong) had been responded to differently on a second occasion (e.g. making it also an Old-correct trial type).

Initially a within group comparison was made for the control task models between Old-correct and Old-wrong scores. A Wilcoxon matched paired test showed no significant difference between these scores (Z= -1.09, p = 0.27). Therefore subjects' scores for these two question types were summed together and counted under the Old question type. This meant that between groups comparisons on the Old question type could now be performed with the control task models. The resulting mean percentage of correct responses are shown in Figure 2.2. Due to some of the data being discarded, these results are shown as percentages.

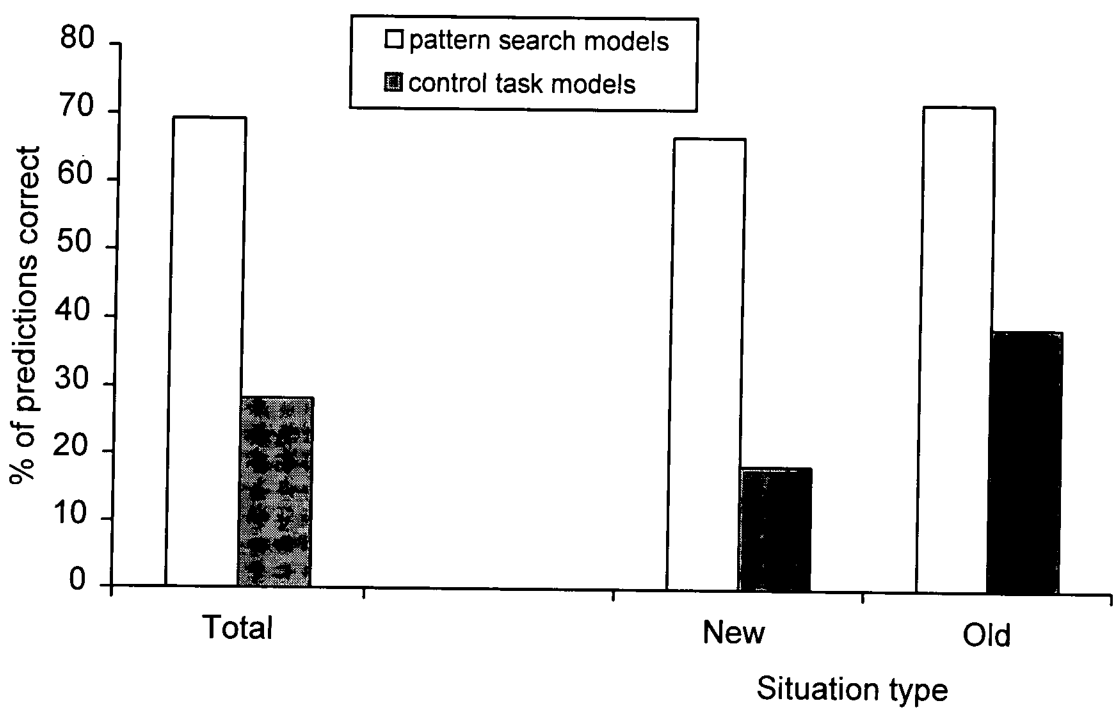


Figure 2.2: Mean percentage of correct responses to each category of prediction question for the models.

The data in Figure 2.2 were analysed using a 2 (learning goal) by 2 (question type: Old vs New) analysis of variance with repeated measures on the last factor revealing a main effect of learning goal, $F(1,22) = 19.26$, $p < 0.001$, a main effect of question type, $F(1,22) = 10.05$, $p = 0.004$, and a significant interaction between learning goal and question type, $F(1,22) = 4.75$, $p = 0.04$. Paired sample t-tests showed a significant difference between Old and New scores for the control task models, $t(11) = 3.05$, $p = 0.01$, but not for the pattern search models, $t(11) = 1.03$, $p = 0.324$. Between group comparisons showed the pattern search models significantly outperforming the control task models for the Old questions, $F(1,22) = 5.95$, $p < 0.03$, and for the New questions, $F(1,22) = 29.55$, $p < 0.001$. For the pattern search models the total prediction question scores and the scores for the two question types were significantly above that expected by chance (scores of 24%, all p values < 0.001). For the control task models the total prediction question score was not significantly different from that expected by chance ($p > 0.1$), however the Old question score was significantly above chance, ($p < 0.05$), while the New question score was not ($p > 0.1$). Thus, the total score was the same as that expected by chance because of the poor performance on the New questions only.

So, for the prediction questions it is clear from the results that the pattern search models outperformed the control task models overall, and for both question types. The pattern search models performed comparably whether subjects were predicting from familiar or novel situations. The control task models however, predicted better from familiar situations. For novel situations the control task models predicted no better than if they had been guessing.

To examine how much the predictions relied upon control performance, total prediction scores were correlated with the number of correct trials during the test phase. Spearman rank correlation coefficients were 0.10 ($p > 0.7$) for the control task models and 0.67 ($p < 0.02$) for the pattern search models. The control task models' prediction scores were also correlated with the number of correct trials during the *learning* phase. The correlation coefficient was 0.21 ($p > 0.5$). As can be seen the correlations were significant only for the pattern search models. This suggests that only for the pattern search models was test performance directly related to prediction performance. As expected however, for the control task models it appears that how well they did on the predictions bore no strong relation to how they did on controlling Clegg.

The Rule Description Questions:

The answers were judged by two judges and ranked into three categories; *No information or Wrong, Partially Correct, Correct.* The ranking procedure was identical to that used in Experiment 1 (see pg. 32). Both judges ranked the answers identically. These rankings can be seen in Table 2.2.

Table 2.2
Ranking of answers to the rule description question.

	Category	No information or Wrong	Partially Correct	Correct
Performance of Models				
	Model's goal			
	Control Task	**11**	**1**	**0**
	Pattern Search	**2**	**3**	**7**
Performance of Observers				
Model's goal	*Observer's goal*			
Control Task	Pattern Search	**4**	**4**	**4**
	Control Task	**10**	**1**	**1**
Pattern Search	Pattern Search	**3**	**1**	**8**
	Control Task	**10**	**2**	**0**

Fisher exact probability tests were performed comparing the number of answers in the *No information or wrong* category and in the *Correct* category. This showed that the pattern search models performed significantly differently from the control task models, ($p < 0.01$). It is clear from Table 2.2 that the control task models had more answers ranked in the *No information or wrong* category than the pattern search models. It is also clear that the pattern search models had more answers ranked *Correct*. Thus, the pattern search models were significantly better than the control task models at producing answers that contained declarative knowledge. As suggested in Experiment 1, it might be argued however, that too strict a criterion was used to categorise answers as *Correct* and that with a looser criterion the two sets of groups may have been more similar. With this possibility in mind another Fisher exact probability test was carried out, but this time the number of questions in the *Partially Correct* and *Correct* categories were added together. The tests showed the same result as with the stricter criterion ($p < 0.01$).

The Range Of Interactions

To explore how the range of interactions differed between the two model groups, how many times the computer or the model used a particular response during the learning phase was examined, i.e. the amount of times each state was encountered was examined. The average number of times a response was used was then calculated for each group. These can be seen in Figure 2.3.

Independent sample t-tests showed that for 7 out of the 12 responses the groups differed at a statistically significant level (all p values < 0.05). For the first three responses (*very rude, rude* and *very cool*) the pattern search models had a significantly higher average use. For three of the middle responses (*polite, very polite* and *friendly*) the control task models had, on average, more encounters. For the last response on the scale (*loving*) the pattern search models had more encounters.

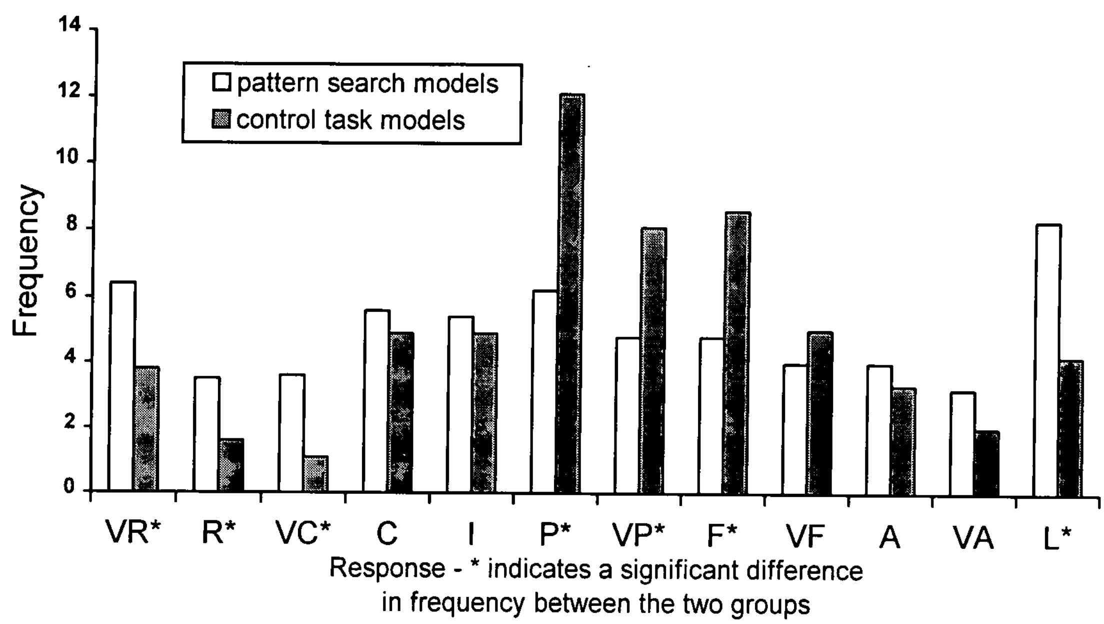


Figure 2.3: State encounters for the two model groups.

The above pattern occurred as the control task models' encounters centred around the target goal of *polite* and the encounters with the other responses dropped off each side of this target. However, as expected the pattern search models' encounters were approximately even across the scale and hence at the edges of the scale this group was encountering significantly more responses than the control task models. Also calculated was the standard deviation of the average number of times any response was used by each group. For the control task models the average standard deviation was 4.3 and for the pattern search models, 3.3. As expected, this shows that the pattern search models encountered the different behaviours in a more even way, whereas the control task models had more interactions with certain responses over others. Additionally, how many different state transitions the two model groups encountered during the learning phase was explored. There was a maximum number of 30 possible state transitions that each group could encounter during the learning phase (one for each trial). The control task models encountered an average of 20.6 state transitions and the pattern search models encountered 23.1 . An independent sample t-test showed a significant difference between the two groups for these two values $t(22) = -2.11$, $p < 0.05$. This shows that as predicted the pattern search models encountered a significantly larger amount of state transitions than the control task models.

The Observers

To make the results section for the observers clearer, comparisons to chance for all measures are presented here at the beginning of this section. Scores expected by chance were 7.4 for the test phase total scores, and 24% for the prediction questions total scores, and for the New and Old question scores. For each set of observers their scores for these measures were compared to that expected by chance. Both sets of pattern search observers, regardless of the goal of their model scored above chance on all measures (all p values < 0.02). Neither set of control task observers were significantly different from that expected by chance on any measure (all p values > 0.3).

Test Trials

The mean number of correct trials, for the entire test phase and each half of the test phase can be seen in Figure 2.4.

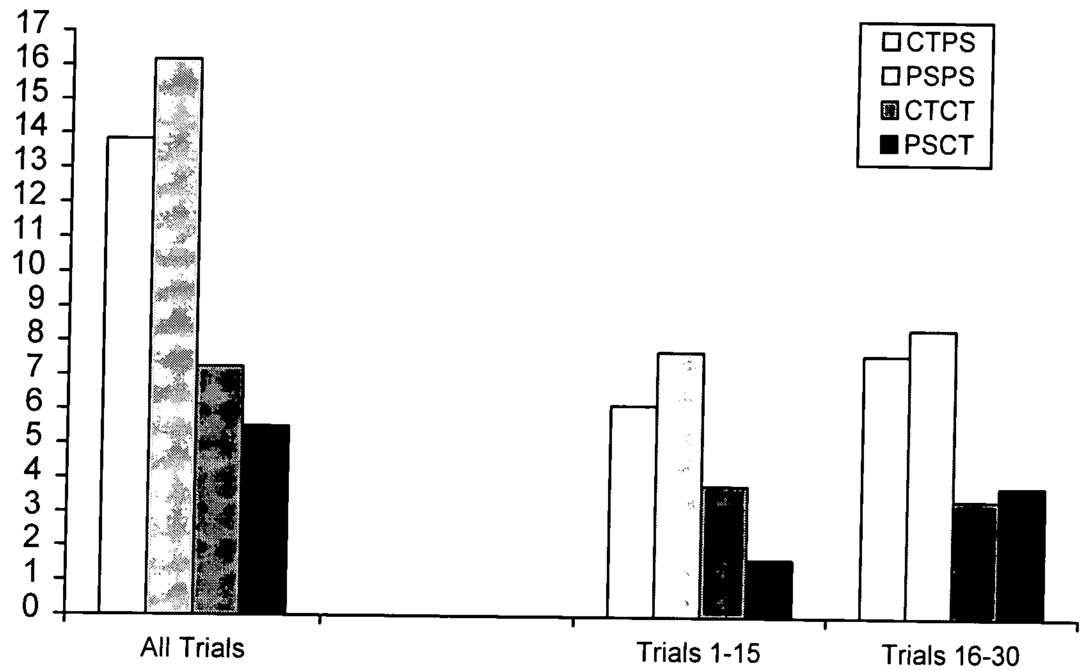


Figure 2.4: Mean number of correct trials in the test phase for the sets of subjects in the observing groups. Data are shown for all 30 trials combined and for the first and second 15 trials. (Note: In the legend, the first 2 letters identify the goal of the model, the second 2 identify the goal of the observer. CT = Control Task; PS = Pattern Search.)

The data in the Figure were analysed using a 2 (goal of model) by 2 (goal of observer) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results showed a main effect of goal of observer, $F(1,44) = 23.37$, $p < 0.001$: the pattern search observers outperformed the control task observers overall on the test trials. There was also a main effect of trial block, $F(1,44) = 5.5$, $p = 0.024$: subjects performed better in the last 15 trials than in the first 15. There was however a significant three-way interaction between trial block, goal of model and goal of observers, $F(1,44) = 4.18$, $p = 0.047$. There were no other significant interactions or effects. It is clear from these results that the goal the models had been given did not affect the performance of the observers during the test trials. However the goal of the observers did affect the results. The pattern search goal produced significantly better performance than the control task goal. All subjects improved during the test phase. Inspection of Figure 2.4 indicates that the three way interaction seems to be due to PSCT observers performing poorly on the first half of the test phase.

The Prediction Questions

The mean percentage of correct responses for the total, Old and New question types are shown in Figure 2.5.

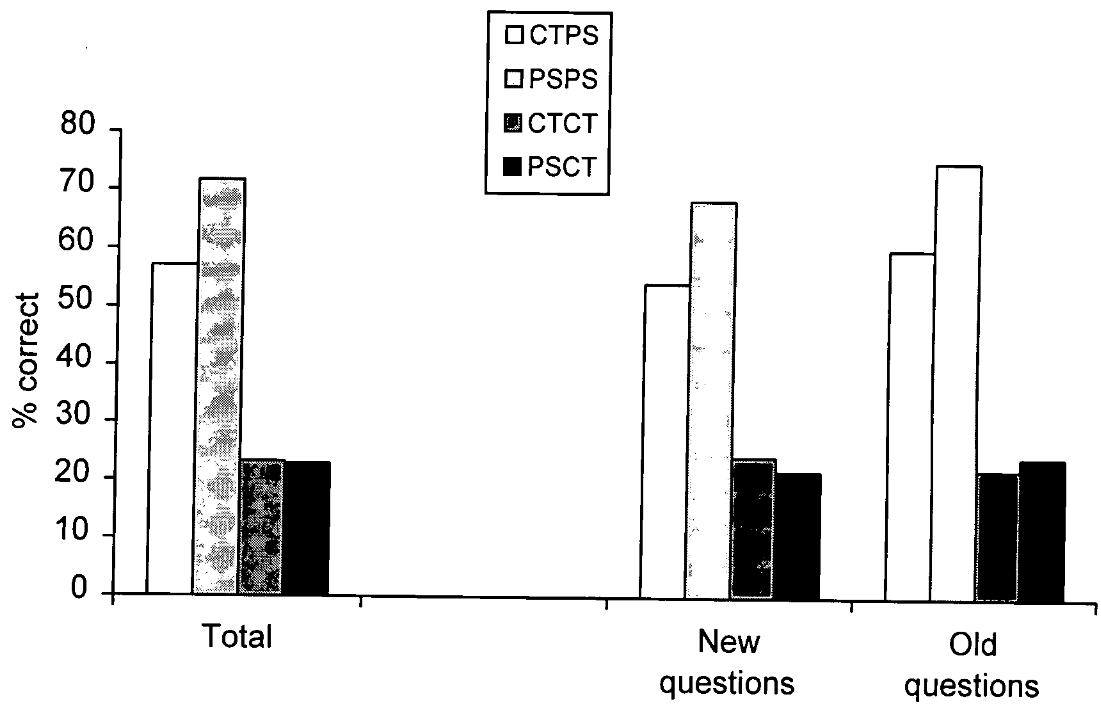


Figure 2.5: Mean percentage of correct responses to each category of prediction question for the sets of subjects in the observing groups.

The data in the Figure are analysed be a 2 (goal of model) by 2 (goal of observer) by 2 (question type) analysis of variance with repeated measures on the last factor. The results showed a main effect of goal of observer, $F(1,44) = 47.59$, $p < 0.001$: the pattern search observers outperformed the control task observers overall on the prediction questions. There were no other main effects and no significant interactions. These results again suggest that the goal of the models had no influence on the observers' performance. The two different sorts of prediction question were also answered comparably. The lack of interaction between question type and goal of observer suggests that this was the case for both the pattern search observers and control task observers. In the case of the two sets of pattern search observers this was because subjects did equally well on New and Old questions. As detailed at the beginning of the Observers' result section, the two sets of control task observers did not perform above chance on either New or Old questions. Hence, their similarity in performance on the Old and New scores was because observers performed equally *poorly* on New and Old question types.

It was predicted that the pattern search observers would learn explicit rules so perform comparably to the pattern search models. It was also predicted that the pattern search observers would outperform the control task models (predicted to learn instances) on the New questions. Individual comparisons confirmed all these predictions - seen Table 2.3.

Table 2.3
Between groups comparisons for the prediction questions

(1) = control task models; (2) = pattern search models; (3) = CTPS; (4) = PSPS; (5) = CTCT; (6) = PSCT

	(1) New	(1) Old	(2) New	(2) Old	(3) New	(3) Old	(4) New	(4) Old	(5) New	(5) Old
(1)										
(2)	29.5 .001	5.95 .03								
(3)	10.4 .004	1.43 ns	0.9 ns	0.9 ns						
(4)	36.9 .001	8.59 .008	0.02 ns	0.1 ns	1.2 ns	1.6 ns				
(5)	1.2 ns	5.6 .03	23.5 .001	33.9 .001	7.43 .02	13.4 .002	29.8 .001	48.4 .001		
(6)	0.39 ns	5.95 .03	26.1 .001	29.6 .001	8.68 .008	11.7 .003	32.9 .001	41.6 .001	0.2 ns	0.09 ns

(6) = PSCT ns = not significant F(1,22) value
Sig. *p* value

There were no significant differences between the two sets of pattern search observers and the pattern search models on either Old or New questions (see column (2) rows (3) & (4) of Table 2.3), and the two sets of pattern search observers outperformed the control task models on New questions (see column (1), rows (3) & (4) of Table 2.3). It was also predicted that the control task observers may learn instances and so underperform compared to pattern search observers and models. These predictions were also confirmed, as can be seen by inspecting rows (5) & (6), column (2), (3) & (4) of Table 2.3. Finally, it was suggested that if the control task observers had learnt instances they should perform comparably to control task models, while if they failed to learn at all, they would underperform compared to control task models. The results of the individual comparisons supported the latter prediction. Recall that for the Old questions the control task models performed above chance and the control task observers performed comparably to chance. It was on these scores that the control task models outperformed the control task observers (see column (1), rows (5) & (6) of Table 2.3).

Total prediction scores were correlated with the number of correct trials during the test phase. Spearman rank correlation coefficients were 0.72 ($p < 0.009$) for the CTPS observers, 0.73 ($p < 0.007$) for the PSPS observers, -0.04 ($p > 0.9$) for the CTCT observers and -0.26 ($p > 0.48$) for the PSCT observers. Thus, the correlations were significant for the two sets of pattern search observers. This suggests that for these sets of subjects, as for the pattern search models, performance during the test phase was directly related to prediction performance. As expected however, the prediction performance of the control task observers, like that of the control task models, bore no strong relation to how they did on controlling Clegg.

The Rule Description Questions:

The answers were judged by two judges and ranked into three categories; *No information or Wrong, Partially Correct, Correct.* The ranking procedure was identical to that used for the models. Both judges ranked the answers identically. These rankings can be seen in Table 2.2 (pg. 58).

Fisher exact probability tests compared the number of answers in the *No information or wrong* category and in the *Correct* category. The two sets of pattern search observers performed comparably (p value > 0.2) and significantly better than the two sets of control task observers (all p values < 0.01) who, in turn, performed comparably (all p values > 0.5). As before, it might be argued that too strict a criterion was used to categorise answers as *Correct* and that with a looser criterion the two groups may have been more similar. Fisher exact probability tests were repeated with the number of responses in the *Partially Correct* and *Correct* categories added together. Again the tests showed the same pattern as the one described above.

The Influence Of The Models On The Observer Groups

To examine whether there was any relation between an observer's performance and the model's performance, correlations were conducted on the New and Old prediction questions and on the correct responses on the test trials. Correlations compared the control task models with the CTPS observers and the CTCT observers and they compared the pattern search models with the PSPS observers and the PSCT observers. The Spearman rank correlation coefficients and their significance levels can be seen in Table 2.4. None of the coefficients were significant, suggesting the model's performance had no effect on observer's subsequent performances.

Table 2.4
Spearman Rank Correlations of observer's scores with their model's scores.

Scores	New	Old	Total test phase
Comparisons			
control task models compared to CTPS observers	**0.17** $p > 0.5$	**0.39** $p > 0.2$	**0.44** $p > 0.15$
control task models compared to CTCT observers	**-0.22** $p > 0.5$	**-0.28** $p > 0.35$	**-0.36** $p > 0.25$
pattern search models compared to PSPS observers	**0.09** $p > 0.7$	**0.06** $p > 0.8$	**-0.31** $p > 0.3$
pattern search models compared to PSCT observers	**-0.004** $p > 0.99$	**-0.12** $p > 0.7$	**-0.06** $p > 0.87$

DISCUSSION

The key purpose of this study was to establish why non-specific, pattern search instructions lead to rule learning. One possible reason, the 'salience explanation', is that the lack of a specific goal means that a larger range of interactions are explored and this makes the underlying pattern more salient. The alternative, the 'goal explanation', is that rule learning occurs due to a more fundamental feature of non-specific, pattern search instructions - that is, they allow the subject to leave instance space and explore rule space instead. As predicted, the results from this experiment rule out the salience explanation and thus add more weight to the goal explanation. These conclusions are deduced from a number of key points described below.

Firstly, the data from the control task models and the pattern search models, even in the tougher conditions set by the design, replicated the results from Experiment 1. The pattern search models learnt rules while, as in previous studies (e.g. Berry & Broadbent, 1984; Dienes & Fahey, 1995, Geddes & Stevenson, in press), the control task models learnt instances. As in Experiment 1, that pattern search models learnt explicit rules is deduced from the following points: (1) The similarity of performance on both New and Old prediction questions; (2) The success at answering the rule description questions; (3) The significant positive correlation between total questionnaire score and test phase performance. Evidence for the control task models learning instances is inferred from the following points: (1) The poor performance on the New compared to the Old prediction questions; (2) The poor answers to the rule description questions; (3) The lack of a significant positive correlation between total prediction scores and trial performance. The notion of the two groups learning differently is further reinforced by the comparisons between the two groups; the pattern search models performed better on both the prediction questions and the rule description questions, indicating that they had acquired significantly more verbalisable knowledge than the pattern search models.

Secondly, the results met the prediction of the pattern search models encountering a wider range of interactions than the control task models. The control task models' interactions centred around the target response of *Polite* with responses adjacent to this response declining in frequency. The pattern search models' interactions were roughly spread across all responses. Also, to support this notion of the pattern search models encountering a wider range of interactions, the results showed that these subjects encountered significantly more state transitions than the control task models. This establishes the 'salience explanation' as feasible and hence makes it necessary to eliminate.

Thirdly, when the range of interactions was reduced by asking observers to observe models given control task instructions, the observers can still learn explicitly accessible rules if given pattern search instructions. The results from the CTPS observers lead to this deduction. It is concluded that these subjects learnt explicitly accessible rules because their pattern of data was almost identical to the pattern search models'. Not only did the CTPS observers perform equally well on Old and New prediction questions and have a positive correlation between control performance and prediction question score, but in comparison with the control task models (labelled as undergoing instance learning) they outperformed these models on the New questions, on the prediction questions and on the rule descriptions. However, the PSPS observers also learnt rules. With both these observer groups learning rules there is a chance that the very act of observing leads to rule learning. If this were shown to be so, it would prevent the conclusion being made that rule learning is primarily produced from having pattern search instructions. To rule out the notion that the act of observing is a stimulator to rule learning, situations need to be shown when observing doesn't lead to rule learning.

So, to meet this last requirement; control task observers were included in the experiment. The results showed that the control task observers performed poorly irrespective of how wide a range of interactions they observed. This is deduced from the facts that for the two sets of control task observers (i) both sets underperformed compared to all the other groups on all measures (except the control task models on the New questions), (ii) the performances of neither set differed from chance on any of the measures examined, and, (iii) there were no significant differences between the two sets. Consequently it can be concluded that observing is not a stimulant to explicit rule learning. Instead, rule learning is induced by pattern search instructions. Further, observing, when combined with control goal instructions, appears to be thoroughly detrimental to learning. All in all, the results of observing with control goal instructions match those of Berry (1991), suggesting that the minor differences in methodology have negligible effect on the results.

In summary, the results from Experiment 1 were replicated; the non-specific pattern search goal leads to rule learning whereas the traditional specific control goal lead to instance learning. The results indeed showed that the pattern search models encountered a more varied set of interactions than the control task models, thus setting up the 'salience explanation' as feasible and therefore making it necessary to eliminate. The pattern search observers while observing the narrower range of interactions of the control task models, managed to achieve rule learning and so this rules out the salience explanation. However the pattern search observers, while observing the pattern search models, also learned the rule, so it is possible that the very act of observing leads to rule learning. This clearly however is not the case as control task observers did not learn at all. So, finally this allows the 'salience explanation' to be firmly eliminated and the conclusion can be made that rule learning stems from non-control oriented pattern search instructions for a reason other than the fact that these instructions lead to a wider range of interactions being encountered.

The effect on learning of *observation* combined with pattern search instructions does not appear to be detrimental: There was no notable difference between the CTPS observers, the pattern search models *or* the PSPS observers. The results also suggest that the PSPS observers learned explicitly accessible rules. The lack of difference between the CTPS observers and the PSPS observers goes on to strengthen the argument that the range of interactions subjects see is not vital in dictating the nature of their subsequent learning. This point is further reinforced by examining the correlations between models' and observers' scores. The results showed no significant correlations This suggests again that how an observer performs is not related to how a model performed - despite the fact that they were both viewing an identical range of interactions. However, the lack of correlation between the control task models and the CTPS observers could be explained by the fact that the two groups of subjects had different learning goals - control task and pattern search respectively. No such explanation can be used for the lack of correlation between the pattern search models and the PSPS observers; both groups had the same learning goal. The lack of correlation between these two groups is more evidence that rule learning is not closely connected to the range of interactions that a subject may see.

Pattern search instructions have been shown here and in Experiment 1 to lead to rule learning irrespective of whether or not subjects have to observe during their learning period. Control task instructions have been shown to lead to instance learning only when subjects are not hampered by observations. Therefore, observations can be seen to have no effect on rule learning, but to have a considerable effect on instance learning. The main affect of observing is that it denies subjects the chance to make actions during their learning period. The results in this study then, would suggest that actions appear to be vital for instance learning but not for rule learning. The fact that actions are needed for instance learning is in direct support of Berry's (1991) finding. This finding demonstrates the close link between instance learning and actions and highlights the similarity between instance learning and procedural learning. One alternative proposal is that instance learning may not be prevented by the lack of actions, but by the lack of feedback. Perhaps it is feedback that is vital for instance learning. Observers were not given feedback as they were not told the model's goal.

A potential problem with the experiments in this study is related to the order of the goals given to the control task models between the learning and test phases. The group starts off with a goal of getting Clegg to be '*Polite*' and then has to make Clegg be '*Very Friendly*' during the test phase. It is possible that this order of goals in some ways confounds the results. However in Experiment 1, the order of the control goal in the learning and test trials was reversed for half the subjects in each experimental group. The results showed no difference between the reversed groups and the originals. Thus the specific order of learning and test goals is unlikely to have affected the results of the present experiment.

The alternative proposal for why non-goal oriented, pattern search instructions lead to rule learning was the 'goal explanation'. That is, the pattern search instructions have no element of a specific control goal and so instance learning is not activated. With the 'salience explanation' ruled out, this proposal holds a lot more weight. Leading from the 'goal explanation' is the notion that instance learning in the person interaction task is primarily induced by the presence of a specific control goal rather than by the lack of salience of the pattern. This idea has important implications for the description of instance learning and parallels with procedural learning spring to mind. What is called for now is a direct examination of the nature of the cognitive processes that are stimulated by the different learning goals.

Chapter III

Experiment 3: Learning Goals and Explanations - How Do They Influence Cognitive Processes?

Experiment 1 showed that the goal of a task plays a fundamental role in the type of learning that results from that goal. Experiment 2 partly explored how different learning goals influence learning. One suggestion was that different learning goals indirectly influence learning as the different goals alter the range of information learners experience. The different quantities of information subjects experience may effectively alter the salience of the task for the different goals and perhaps this is how learning goals influence learning. The results from Experiment 2 ruled out this proposition. The alternative suggestion was that learning goals directly influence cognitive processes. In terms of the dual space model, it was suggested that learning goals influence cognitive activities by directing attention to one or both of rule space or instance space. The study reported here directly examines this proposition. The second aim of this study is to examine the impact of self-explanations on learning. The idea being that explanations should enhance learning over and above the level of learning observed in Experiment 1.

To test the proposition that different cognitive processes are induced by different learning goals, subjects learned the 'Clegg' version of the person interaction task, but carried out a concurrent secondary task at the same time. The secondary task required the subjects to talk aloud while learning, either explaining why they were doing what they were doing or describing what they were doing. The results showed that giving descriptions was compatible with the learning processes employed by subjects with a control task goal, whereas giving explanations was compatible with the learning processes employed by subjects with a pattern search goal. This confirmed the notions regarding the hypothesised cognitive processes that each goal induces. The control goal causes subjects to focus on instance space which is compatible with subjects describing what they are doing. The pattern search group causes subjects to have a greater focus on rule space which is compatible with subjects explaining why they are doing what they are doing. To examine the impact of explanations on learning, the focus was on subjects whose secondary task was explaining what they were doing. By making comparisons with the silent subjects' data from experiment 1 and the describing subjects from this study, the results confirmed that explanations do indeed enhance learning, particularly if combined with a learning goal that encourages the exploration of rule space. Finally, the results of this study also support the notion that a dual goal, consisting of both the specific and the pattern search goals, causes a cognitive load that is too heavy to allow either form of learning that would normally result from pursuing one of these goals in isolation.

This introduction is organised as follows: First, the dynamic systems studies that get subjects to concurrently verbalise are examined. Then, there is a discussion as to why self-explanations should enhance learning and the relevant literature on explanations learning is covered. Following this, extra detail is given as to exactly how the secondary tasks used in this study allow the exploration and confirmation of the proposed cognitive processes that different learning goals induce. Finally the hypotheses for the experiment are presented.

The effect of concurrent verbalisation on the 'Clegg' version of the person interaction task has been examined before. Berry & Broadbent's (1984) landmark paper explored the effects of concurrent verbalisation on informed subjects (see Berry & Broadbent, 1984, experiment 3), that is subjects who had the pattern of the underlying task explained to them, and also on uninformed subjects, those that were given no such information. The verbalisation instructions were not as directed as the ones given in the study reported here. Subjects were simply told to think aloud, and were asked to give reasons for their actions. Berry & Broadbent's informed subjects benefited from verbalising and their verbalisations were rule testing in nature. Their uninformed subjects, however, did not benefit from verbalising and these subjects' verbalisations were more general (and obviously less useful) in nature. The sorts of verbalisations coming from the informed and uninformed subjects loosely match the sorts of verbalisations expected to come from the 'explain' and 'describe' verbalisation conditions respectively, used in this study. Therefore, the results of Berry & Broadbent's study can be taken to add support to one of the predictions (detailed below) made in this study, namely, that the 'explain' condition should enhance performance.

Stanley et al (1989) also studied the effect of verbalisation on subjects doing the person interaction task. The required verbalisations were very different from the ones required here. Subjects were told to speak into a tape recorder not continuously, but after every 10 trials. They could say what ever they wanted into the tape recorder, but were coaxed by being told that their verbalisations would be used by an unseen partner at a later date to guide him/her through the task. By making comparisons to silent subjects, the results showed that the very act of verbalising benefited performance. To some degree, the verbalisations were explanatory in nature as subjects had to explain to a second party what was happening in the task. Therefore the verbalisation condition was similar to the 'explain' condition used in this study. From this point of view Stanley et al's result can be taken to add further backing to the prediction that the 'explain' secondary task used in this study should benefit subjects' learning and performance.

Vast methodological differences between both Berry & Broadbent's (1984) and Stanley et al's (1989) studies and the study presented in this chapter mean that there is plenty of originality provided by this study. The prescribed verbalisations in this study are much more directed than in either of the other studies. Also, both the other studies had subjects trying to achieve a specific control goal, whereas the subjects in this study have either a control goal or a pattern search goal or a combination of the two. Additionally, the main focus of this study is not to examine the effects of verbalisations per se, but to use verbalisations to explore the underlying cognitive effect of different learning goals. As a secondary focus of this study however, the effect of verbalisations are directly examined. The 'explain' condition is focused upon to examine the effects of explanations on control goal, pattern search goal and dual goal subjects.

By exploring the effects of explanations on learning, this study specifically examines the effect of self-explanations on learning, not given explanations such as Berry and Broadbent's (1984) informed subjects had. Much of the work on explanation learning has looked at the effect of self-explanations, in which subjects explain to themselves why events are occurring and why they are attempting certain paths of action. They have been shown to aid learning and related performance in a number of studies (e.g. for improvement in; physics learning - Chi, Bassok, Lewis, Reimann, & Glaser, 1989; computer programming - Pirolli & Bielaczyc, 1989; Pirolli & Recker, 1994; general text comprehension - Chi, de Leeuw, Chiu, & LaVancher, 1994).

For example, Chi and Bassok et al studied self-explanations by focusing on the differences in performance of good and poor learners. They used subjects who were part of a longitudinal study of students studying Newtonian mechanics. They gave subjects worked-out problems dealing with the application of Newton's laws of motion. Subjects had to make self explanations aloud while they studied the problems. Good and poor learners were identified through their ability to solve problems based on earlier learned material. The results showed that good learners used self-explanations, poor learners did not. This enabled the good learners to notice comprehension failures more than poor learners, to identify the source of the comprehension failure better than the poor learners, and so to readily locate the information in the previous text that was needed to resolve the comprehension failure.

As another example of the benefits of self-explanation, Chi and de Leeuw et al designed a study whose main purpose was to study the effect of self-explanations irrespective of the individual subject's abilities. They gave subjects some text consisting of 101 sentences on the body's circulatory system. The self-explanation subjects had to read the text and make self-explanations after each sentence and at key points in the text. They were also encouraged to make more detailed explanations if they were not doing so. The control subjects simply had to read the text twice (so that they spent the same time on the text as the self-explanation subjects), but did not have to make any self explanations about the text. The results showed the self-explaining students making the most improvement between pre-text and post-text tests. Also, as in Chi and Bassok et al's paper, the subjects who provided many self-explanations performed better than those who provided relatively few numbers of explanations.

The good learners from the Chi and Bassok et al study were coming away from the problem with knowledge that was not problem specific. This is similar to what the pattern search learners from Experiment 1 were doing. They were coming away from the instances presented during the learning and test phase with more than the knowledge of just the instances - they learned the underlying rule of the task. The self-explain conditions in both the Chi and de Leeuw et al paper and the Chi and Bassok et al paper are almost identical to the 'explain' verbalisation condition used in this study. Therefore it is predicted that the 'explain' condition should yield an improvement in learning above and beyond that of the silent subjects of Experiment 1. Notably, the improvement for the 'explainers' should be particularly apparent when their performance is compared to that of the 'describing' subjects as these will be prevented from performing even internal self-explanations. As explained in the full set of predictions discussed below, these predictions are however complicated by the learning goal that subjects are given,

It was noted above that the good learners from the Chi and Bassok et al paper are similar to the pattern search learners from Experiment 1 who were not self-explaining. However despite the already evident similarities, the effect of self-explanations should still improve the pattern search subjects' learning. The performance of the pattern search subjects in Experiment 1 was very good, but it was not optimal. As Vollmeyer, Burns and Holyoak (1996) point out, inducing subjects to engage in hypothesis testing does not necessarily mean that they will use optimal strategies of hypothesis testing. Vollmeyer et al found that instructing subjects in efficient hypothesis testing enhanced performance on a novel specific goal, irrespective of whether the initial learning goal was specific or non-specific. Self-explanations, having been shown to be a powerful enhancer of learning, might also be expected to increase the efficiency of hypothesis testing. It was explained in Chi and Bassok et al's study that the good learners were able to guide their self-explanations by "accurate monitoring of their own understanding and misunderstandings (Chi and Bassok et al, 1989, pg. 145)", presumably something akin to 'explainers' being led to monitor the validity of their hypotheses. The expectation in this study is that the self-explanations (induced by the 'explain' secondary task), will make it clear to the learner when he or she does not fully understand the rule and thus motivate a further search of hypothesis space to modify or refine the current hypothesis. Consequently, it is predicted that non-specific, pattern search subjects who give explanations should outperform pattern search subjects with a 'describe' secondary task and also the silent pattern search subjects from Experiment 1. The effects of self-explanations on the other goal groups are discussed below.

For the remainder of this introduction, extra detail shall be given to explaining how the secondary tasks can be used to confirm the proposed interpretation of the cognitive processes that different learning goals induce. Throughout this thesis the dual space model has been used to explain the patterns of results that are caused by different learning goals. It has been suggested that a control goal leads subjects to explore instance space and a pattern search goal results in primarily a search of rule space. The secondary tasks presented in this study have been specifically designed to help the subjects focus upon either rule space or instance space. Consequently, this allows one to hypothesise that certain combinations of learning goal and secondary task should be compatible and other combinations should be incompatible. With these hypotheses met there can be stronger grounds for using the dual space model to describe the learning of the different goal groups. It also will allow a closer scrutiny of what processes are involved in the secondary tasks and therefore come to a closer understanding of the processes induced by the learning goals.

For subjects with a control goal it has been postulated that the control task confines and encourages learners to explore instance space. The 'describe' secondary task is specifically designed to confine and encourage learners to explore instance space. By purely describing each and every action, subjects will be focusing on the surface features of the situation and therefore on instances and instance space. The constant demand to make descriptions will completely prevent subjects from hypothesising about the general underlying abstract rule of the task and attempting to explore the other space, rule space. If control goal subjects do not have their learning interfered with by the 'describe' secondary task it will support the theory that it is the control goal that prevents rule learning, with the control goal leading to instance learning because it encourages the exploration of instance space only. For, lack of interference will suggest that the two tasks share the same processes. When subjects are learning instances they are essentially focusing upon each instance they encounter and assessing whether their present actions have led them to acquire or maintain the required goal. The secondary task of describing should be compatible with the 'focusing upon instances' part of the processes a control goal encourages. In other words, describing aloud is simply focusing on instance space out loud as opposed to silently focusing on instance space. Predictably therefore, there should be no conflict of cognitive processes between a secondary task of describing and a specific learning goal. Also, the 'describe' secondary task should not interfere with their normal pattern of learning, i.e. that expected from silent subjects.

For subjects with a non-specific, pattern search goal it has been postulated that the pattern search goal encourages learners to explore rule space. The 'explain' secondary task is specifically designed to encourage learners to explore rule space. By having to explain one's actions, subjects will tend to hypothesise about why they have reached their present situation and what action their next input of 'x' is likely to have. In other words the 'explain' task will give subjects a greater tendency to hypothesise and thus explore rule space. If the pattern search group do not have their performance interfered with by the 'explain' task it will support the notion that the pattern search goal leads to explicit rule learning as it encourages subjects to explore rule space. The lack of interference will suggest that the two tasks share the same processes. It is suggested that there should be more than lack of interference. Bearing in mind the work on explanation learning, it is predicted that explaining aloud should enhance the learning of pattern search goal subjects by encouraging them to continuously review their hypotheses. It should be noted that explanations, though predicted to enhance learning, are only predicted to improve it not change the nature of it. In other words, subjects should still learn rules and therefore, their pattern of learning should still be the same as that of the silent subjects.

In the last two paragraphs, the compatible conditions of learning goal and secondary task have been considered. It has been predicted that a 'describe' concurrent task should be compatible with a specific, control goal. It has also been suggested that an 'explain' concurrent task should be compatible with a non-specific pattern search goal. Indeed, it was predicted that that particular combination should enhance performance. The incompatible mixes of secondary task and learning goals are now considered.

For the subjects with a control goal doing an 'explain' secondary task a drop in performance is not actually expected. The 'explain' secondary task does not have to conflict with the control goal. It is probable that subjects will simply explain how they are trying to reach their control goal. This should lead them to vigorously use means ends analysis to reach their goal (what the dual space model indicates they should be doing while exploring instance space). Therefore the 'explain' task will not necessarily make them explore rule space. The more vigorous use of means ends analysis may actually lead to some improvement in their performance. However, as subjects are confined to instance space by constantly having to achieve a control goal it is unlikely that the 'explain' task will lead to explicit rule learning of the underlying rule that controls Clegg. So, their general pattern of learning should stay the same as that expected from silent or 'describing' subjects.

For the subjects with a pattern search goal doing a 'describe' task, a considerable impact on performance would be expected. The secondary task will occupy their working memory and prevent the subjects from using it to form hypotheses. That is, giving descriptions of what they are doing should interfere with these subjects' ability to generate and test hypotheses. Therefore explicit rule learning is unlikely and a general performance drop is probable. Also the normal pattern of learning (i.e. that expected from silent pattern search subjects) would not be expected.

Subjects with a dual goal have a distinctive pattern of learning that is distinguishable from both control task and pattern search subjects. It has been suggested in chapter 1 (and will be investigated and confirmed in Experiment 4) that this group of subjects learns instances implicitly as their dual goal overloads working memory. The main reason for the working memory overload is that the two goals conflict with each other and prevent either goal from having enough mental resources to lead to a successful result. For the study presented here, it is suggested that a secondary task will reduce the effect of resources being evenly split between the two goals. The secondary tasks should encourage more resources to be guided towards one particular goal. The 'describe' task should encourage dual goal subjects to focus more on the control goal, whereas the 'explain' task will encourage more resources to go towards the pattern search goal. In effect, by allowing more resources to go to achieving one goal or the other, both secondary tasks should improve performance. There should then be enough resources to actually make some headway with which ever secondary task has the compatible learning goal.

Specifically the hypotheses are; For dual goal subjects with both an 'explain' and 'describe' secondary task there should be some improvement over performance of subjects without a secondary task (i.e. the silent subjects from Experiment 1). How their patterns of performance compares to the silent subjects should give some indication of how much impact on learning the verbalisation conditions cause.

For subjects with a pattern search goal and an 'explain' secondary task, it would be expected that subjects will learn explicit rules and outperform the silent subjects from Experiment 1 and the 'describing' pattern search subjects from this study. However the general pattern of results as seen for the silent subjects should still be displayed as subjects are predicted to learn better not differently. For the subjects with a pattern search goal and a 'describe' secondary task it would be expected that the normal rule learning of the pattern search subjects will be impaired. Performance, therefore should be below that of both silent subjects and 'explaining' subjects. Also, their pattern of performance should be markedly different from that normally expected for pattern search subjects and hence that seen for the silent subjects and that expected for the 'explaining' pattern search subjects. As they should no longer be learning rules, they should be particularly poor on New prediction questions (those prediction questions that require subjects to know the underlying rule of the task). Exactly how the 'describing' pattern search subjects' pattern of performance compares with other groups in this study should give some indication of how these subjects learn.

For the control goal subjects with a 'describe' secondary task it would be expected that this compatible condition will lead to performance that is comparable to silent control goal subjects. For the control task subjects with an 'explain' secondary task, it would not be expected that performance will be below that of silent subjects as might be predicted for an incompatible secondary task. There may instead be some improvement as the 'explain' task may lead subjects to have a greater focus of concentration on their control task. However their general pattern of learning should remain the same.

In the specific predictions made above, there have been references to patterns of performance varying or staying the same as those of the silent subjects from Experiment 1. The main defining characteristic of a pattern of learning comes from the prediction questions and subjects' relative performance on the different prediction question types. So, in the specific predictions made above, references to patterns of performance varying or staying the same will be examined by looking at the normal pattern of prediction questions displayed by the silent subjects from Experiment 1 and how the verbalising groups' patterns of prediction questions compares to them. As a recap of the normal pattern of prediction questions for the different goal groups, remember that the pattern search subjects from Experiment 1 performed comparably on all prediction question types; Experiment 1's control task subjects performed comparably on Old-wrong and Old-correct prediction questions and performed better on these two types of prediction questions than on the New prediction questions; The dual goal subjects performed better on the Old-correct prediction compared to both New and Old-wrong prediction questions which were performed comparably.

METHOD

Subjects*:* The 72 volunteer subjects were Durham University graduate and undergraduate students and some members of the public, aged between 18 and 24.

Design*:* A 2 (verbalisation) by 3 (learning goal) independent groups design was used. Subjects were randomly allocated to one of the three goal groups (control task, pattern search and dual goal group - a combination of control task and pattern search goals). Half the subjects in each goal group were given the 'explain' secondary task and the other half were given the 'describe' secondary task. Subjects with an 'explain' task had to constantly verbalise during the learning and test phases of the experiment *explaining* why they were doing what they were doing. Subjects with a 'describe' secondary task had to constantly verbalise, *describing* what they were doing, not why they were doing it. Subjects were required to complete 30 trials for the learning phase, and another 30 trials for the test phase. They were then given the unexpected questionnaire. This was the same as that described in Experiment 1 (see pg. 23).

The task : Subjects did the Clegg version of the person interaction task which was identical to that described in Experiment 1 (see pg. 21).

For the learning phase the six sets of subjects were given identical instructions except for two sentences. The identical part of the instructions concerned the information describing the nature of the task. Additional to this, subjects were given their appropriate learning goal instructions as described in Experiment 1. Subjects were also given instructions regarding their concurrent verbalisations. Subjects with the 'explain' secondary task were given the following instructions; "Whilst you are doing the task I want you to speak your thoughts aloud. Any thought that comes into your head you must say aloud. I want you to constantly be explaining why you are doing what you are doing. Every action or reaction that is made I want to hear you giving some sort of explanation as to why it was made." Subjects with the 'describe' secondary task were given these instructions; "Whilst you are doing the task I want you to constantly speak aloud. I want you to describe what it is that you are doing, as if you are giving a running commentary of what is happening. Every action you take and response that is given I want you to describe aloud. However, at no point do I want you to explain why you are doing what you are doing." The rest of the experiment was identical for all six sets of subjects and the same as that described in Experiment 1.

The Questionnaire : this consisted of prediction questions and rule description questions identical to that described in Experiment 1.

Procedure: Subjects were randomly allocated to one of the six sets of subjects. As mentioned above, the six sets of subjects received identical initial instructions apart from two sentences. The first sentence dictated the aim of that particular group for their learning phase of the experiment. The second sentence dictated the nature of the verbalisations that subjects had to make during the learning and test phases. Apart from this and the verbalisation instructions, the remainder of the experiment was identical for all groups:

To begin with the microphone was attached to the subject and the tape recorder was set up. The instructions explaining the nature and aim of the subjects' initial learning task were presented. These were followed by the learning phase of the experiment. On completion of this phase all subjects received instructions describing their new aim for the test phase and then the test phase started. Clegg initiated both learning and test phase by displaying one of the three adjectives centred on *Polite*. Following the test phase the subjects were presented with instructions for the prediction questions. These instructions simply described the nature of this new task and provided an example of the information from which they would have to make a prediction. They also explained that each question was unrelated to the previous one. After completing the prediction section subjects were given a pen and paper and were asked to answer the two rule description questions.

Throughout the experiment, all instructions appeared on the computer screen but were also read out to the subjects. The experimenter stayed with the subject throughout the experiment in order to answer any arising questions.

RESULTS

Performance during learning and testing

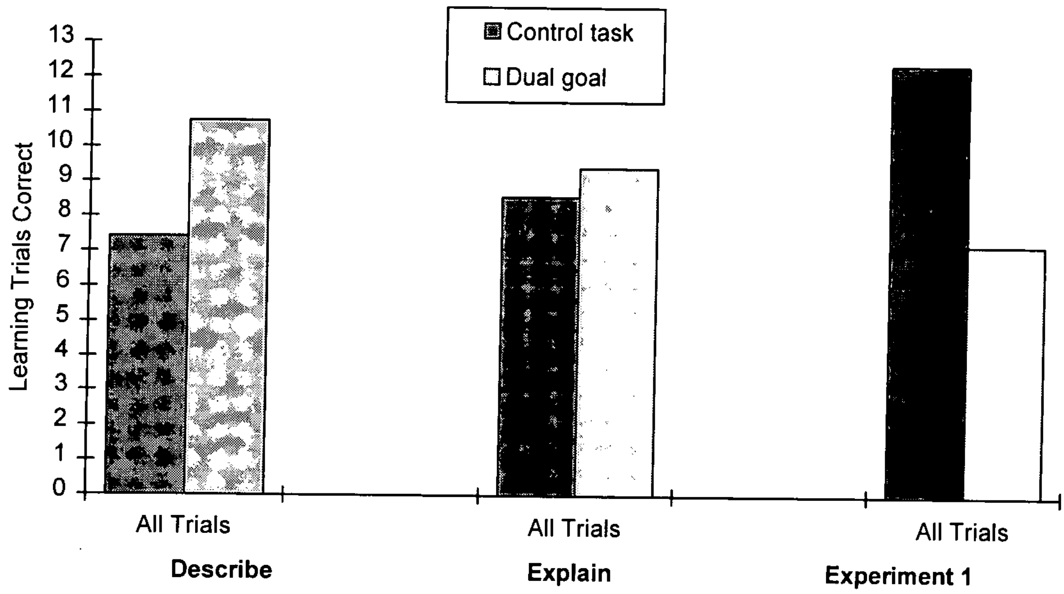
Trials were scored as correct in the same way as was described in Experiment 1 (see pg. 25). Due to the lack of specific aim for the pattern search subjects during their learning phase, no measure could be made for their performance during the first set of trials.

The total number of correct trials during each phase of the experiment and each half of an experimental phase were measured. The mean numbers of correct trials were calculated for each set of subjects in each of these categories. At the end of the results, comparisons are made to the silent subjects from Experiment 1. Therefore, to make the comparisons easier to examine, also included in the Figures are the data from Experiment 1's subjects.

Learning Trials

The mean number of correct trials for each group, in the above mentioned categories for the learning phase can be seen in Figure 3.1.

The data from the bottom chart in Figure 3.1 are analysed using a 2 (learning goal) by 2 (verbalisation: describe vs explain) by 2 (trial block) analysis of variance with repeated measures on the last factor. The only significant result was an interaction between trial block and verbalisation, $F(1,44) = 4.86$, $p = 0.033$.



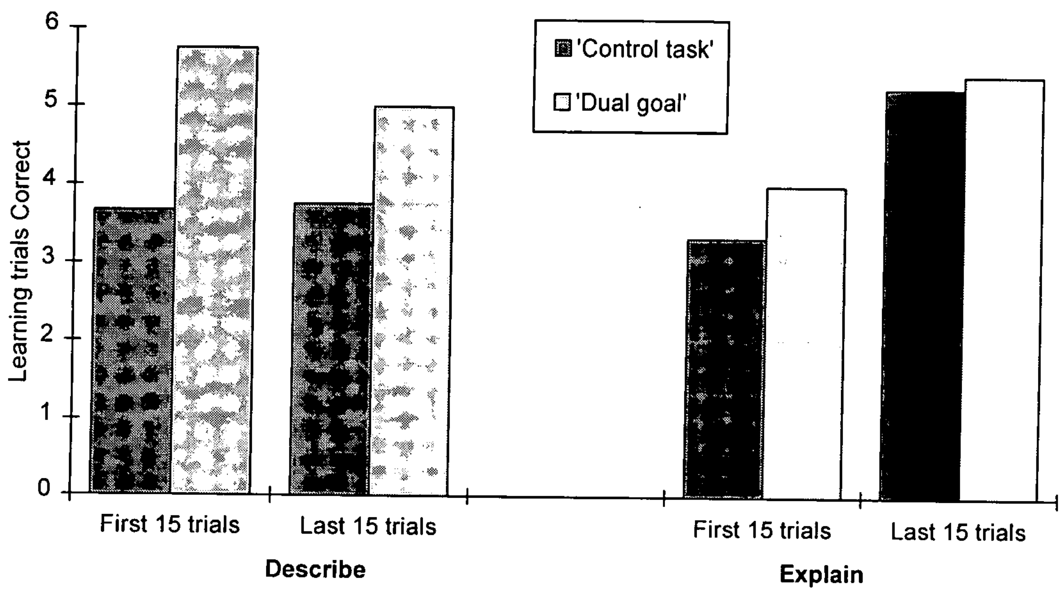


Figure 3.1: Mean number of correct trials in the learning phase for the control task and dual goal groups. Data are shown for all 30 trials combined in the top chart and for the first and second 15 trials in the bottom chart. Experiment 1's data is included for the total phase scores.

To explore this interaction data were collapsed over learning goal and a comparison was done between the first and last 15 trials of the learning phase. The results showed no difference between these trials for the subjects that were concurrently describing what they were doing however there was a significant difference for subjects concurrently explaining what they were doing, t(22) = -2.51, p = 0.02. So, overall during the learning phase, describing had no effect on subjects' performance, while for subjects explaining, learning improved from the first to the second half of the learning phase. (See Appendix 3 for the ANOVA tables and full sets of t-tests for this experiment, pg. 236.)

The Test Trials

The mean number of correct trials for each group, for the entire test phase and each half of the test phase can be seen in Figure 3.2.

The data from the bottom chart in Figure 3.2 were analysed using a 3 (learning goal) by 2 (verbalisation) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results showed a main effect of verbalisation, F(1,66) = 12.73, p = 0.001, and of learning goal, F(2,66) = 4.17, p = 0.02, and a significant interaction between the two, F(2,66) = 9.71, p < 0.001. Inspection of Figure 3.2 (examining the top chart), suggests that this interaction arises because the main effect of verbalisation is confined to 'explain' pattern search subjects who performed better than all other subjects. This observation was confirmed by between group comparisons on the overall scores. The 'explain' pattern search group performed significantly better overall during this phase than all the other groups, (when compared with the 'explain' control task group, F(1,22)=31.65, p < 0.001, 'explain' dual goal group, F(1,22)=13.07, p < 0.001, 'describe' pattern search group, F(1,22)=39.07, p < 0.0001, 'describe' control task group, F(1,22)=24.76, p < 0.0001, 'describe' dual goal group, F(1,22) = 53.95, p < 0.0001). Comparisons between the other groups showed that none of them differed significantly from each other. The effect of trial block just failed to reach significance, F(1,66) = 3.08, p = 0.084: there was a tendency for subjects to get more trials correct in the second 15 trials than in the first. There was also a significant three way interaction between trial block, verbalisation and learning goal, F(2,66) = 3.4, p = 0.039.

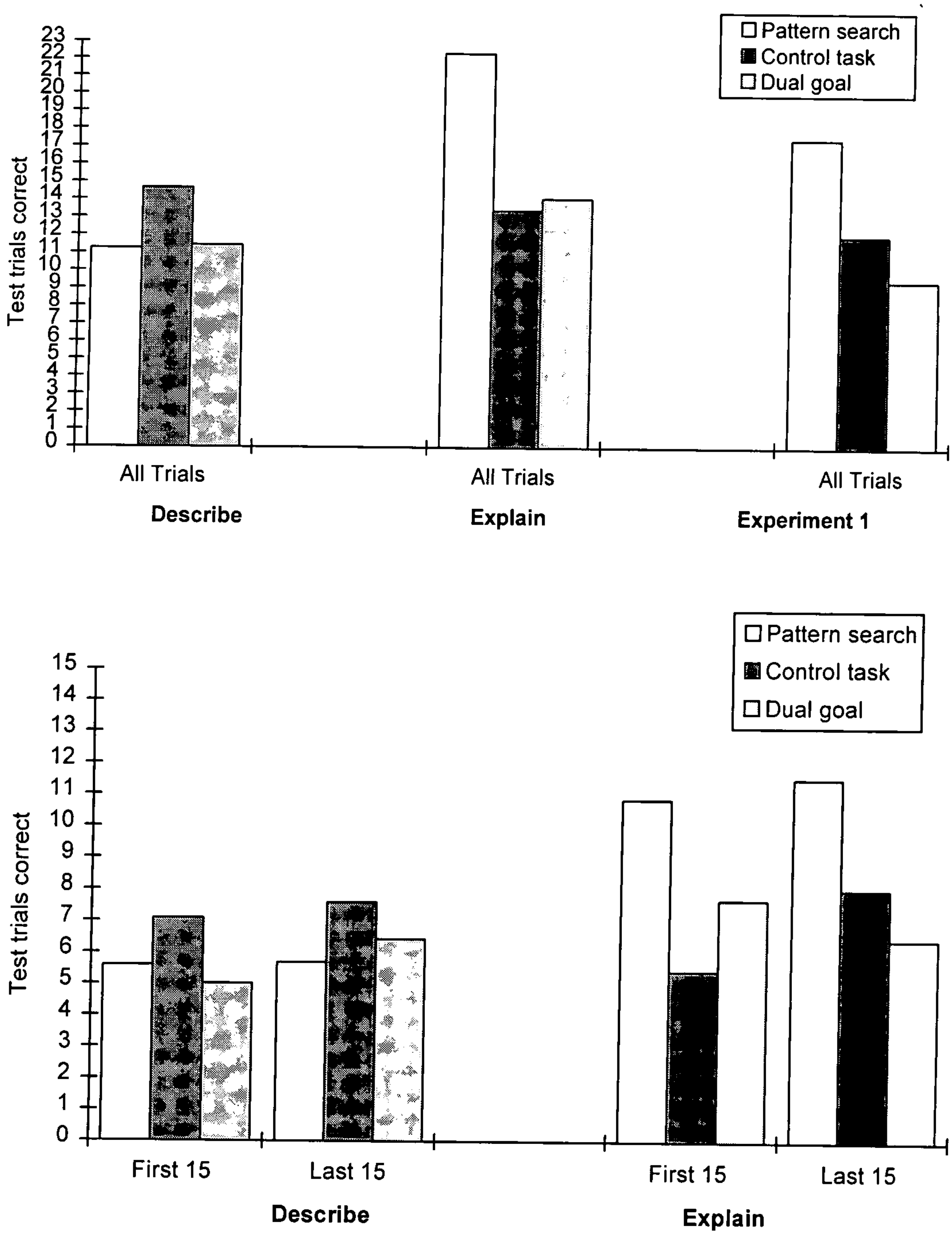


Figure 3.2: Mean number of correct trials in the test phase for all the groups. Data are shown for all 30 trials combined in the top chart and for the first and second 15 trials in the bottom chart. Experiment 1's data is included for the total phase scores.

To explore this interaction, within group comparisons comparing the first and last 15 trials were made for each set of subjects. The only set of subjects to show a significant difference for this comparisons was the 'explain' control task group. This group performed better in the second 15 trials than in the first, ($t(11) = -2.82$, $p < 0.018$).

Transfer: In keeping with the rest of the thesis, another important aspect of performance that was examined is how the groups with a control goal coped when the subjects switched goals from the learning phase to the test phase. Two comparisons were made to examine this: (i) A comparison between the number of correct trials during the last *half* of the learning phase and the first *half* of the test phase and, (ii) A comparison between the *total* number of correct trials during the learning and test phases.

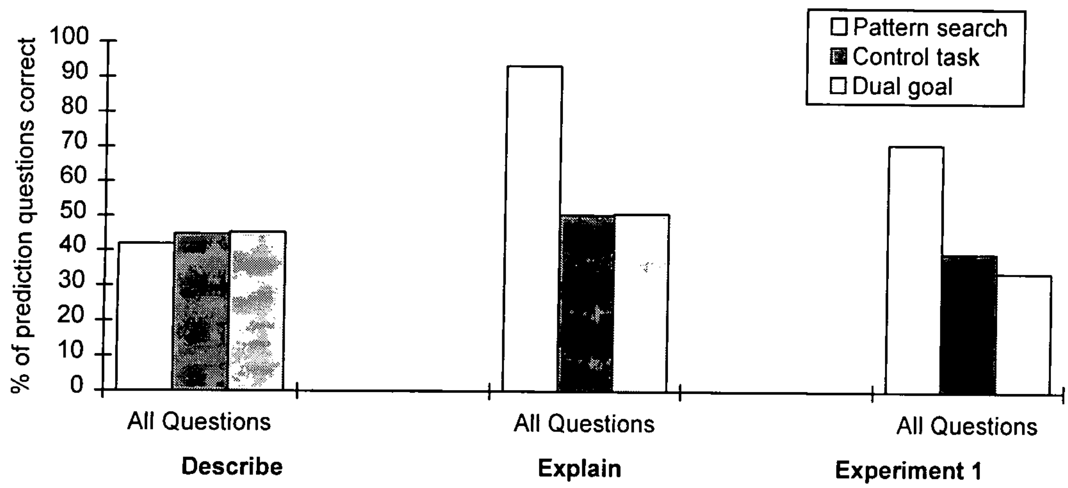
For comparison (i) a 2 (learning goal) by 2 (verbalisation) by 2 (trial block) mixed analysis of variance revealed an effect of trial block $F(1,44) = 10.04$, $p = 0.003$: subjects were performing better during the first half of the test phase than during the last half of the learning phase. The only significant interaction with trial block was the three way interaction between trial block, learning goal and verbalisation, $F(1,44) = 8.91$, $p = 0.005$. There were no other significant main effects or interactions. To explore this interaction within group comparisons compared the last half of the learning phase with the first half of the test phase for each group. Significant differences were shown for the 'describe' control task group, $t(11) = -4.76$, $p < 0.002$, and for the 'explain' dual goal group, $t(11) = -2.36$, $p < 0.04$. In both cases subjects were getting more trials correct during the first half of the test phase.

For comparison (ii) a 2 (learning goal) by 2 (verbalisation) by 2 (phase score) mixed analysis of variance revealed an effect of phase score $F(1,44) = 27.67$, $p < 0.001$: subjects were performing better during the test phase than during the learning phase. However, there was a significant interaction between phase score and learning goal, $F(1,44) = 4.16$, $p = 0.048$ and the three way interaction between phase score, learning goal and verbalisation just failed to reach significance, $F(1,44) = 3.76$, $p = 0.059$. There were no other significant main effects or interactions. To explore these interactions, within group comparisons compared the overall score of the learning phase with that of the test phase for each group. Significant differences were shown for the 'describe' control task group, $t(11) = -4.36$, $p < 0.002$, 'explain' control task group, $t(11) = -2.63$, $p < 0.03$, and the 'explain' dual goal group, $t(11) = -2.89$, $p < 0.016$. In all these cases subjects performed better in the test phase than in the learning phase.

The Prediction Questions

Answers were scored in the same way as in Experiment 1 (see pg. 28). As in Experiment 1, responses were discarded (15% of the data) if a selected trial type (e.g. Old-wrong) had been responded to differently on a second occasion (e.g. making it also an Old-correct trial type). Due to some of the data being discarded, these results are shown as percentages. The mean percentage of correct responses to each question type is shown in Figure 3.3. Due to the refining and discarding of some of the data one of the 'describe' pattern search subjects had no Old-correct type prediction questions and one of the 'explain' pattern search subjects had no Old-wrong prediction questions (Hence the degrees of freedom in the following statistics are adjusted accordingly).

The data from the bottom chart in Figure 3.3 (excluding Experiment 1's data) are analysed using a 3 (learning goal) by 2 (verbalisation) by 3 (question type) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(2,64) = 5.72$, $p = 0.005$, and of verbalisation, $F(1,64) = 22.06$, $p < 0.001$, and a significant interaction between the two, $F(2,64) = 8.5$, $p = 0.001$. To explore these results between groups comparisons were made for the overall prediction scores between all the six groups. The 'explain' pattern search group significantly outperformed every other group (compared with the 'explain' control task group, $F(1,21)=52.11$, $p < 0.0001$, 'explain' dual goal group, $F(1,21)=27.95$, $p < 0.0001$, 'describe' pattern search group, $F(1,21)=26.30$, $p < 0.0001$, 'describe' control task group, $F(1,21)=90.85$, $p < 0.0001$, 'describe' dual goal group, $F(1,21)=48.14$, $p < 0.0001$). Comparisons between the other groups showed they all performed comparably. The analysis of variance also showed a main effect of question type, $F(2,128) = 12.87$, $p < 0.001$, and question type interacted with learning goal, F(4,128) = 3.14, p = 0.017. There were no other significant interactions with question type.



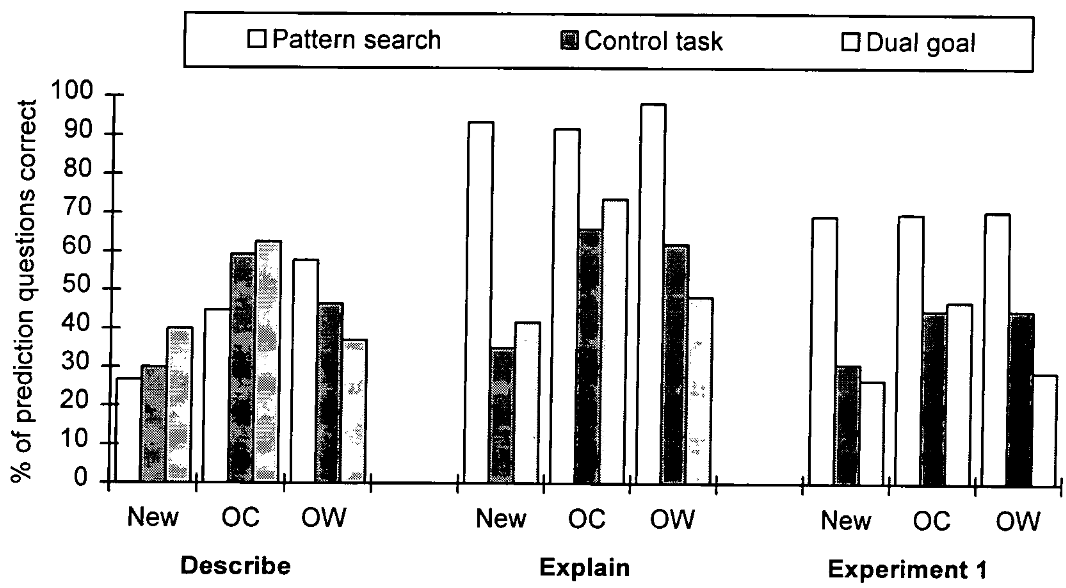


Figure 3.3: Mean percentage of correct responses to each category of prediction questions for each group including Experiment 1's data. The top chart shows data for the overall prediction questions score. The bottom chart shows the data for each question type (OC = Old-correct, OW = Old-wrong).

To explore the interaction between question type and learning goal, within group comparisons were carried out (New vs Old-correct, New vs Old-wrong, Old-correct vs Old-wrong) on each goal group with the data collapsed across verbalisation. The results showed that for the pattern search subjects, performance on the Old-wrong questions was better than on both Old-correct questions ($Z= -2.1$, $p = 0.03$) and New questions ($Z= -2.79$, $p = 0.005$), whilst, the performance on the Old-wrong and New scores was comparable. For the control task subjects performance on the Old-wrong and Old-correct scores was comparable and for both these question types performance was better than on the New questions (Old-correct vs New, $Z= -2.96$, $p = 0.003$, and for Old-wrong vs New, $Z= -2.77$, $p =0.005$). For the dual goal subjects performance on the Old-correct questions was better than on both Old-wrong questions ($Z= -2.6$, $p = 0.008$) and New questions ($Z= -3.06$, $p = 0.002$). The performance on the Old-wrong and New scores was comparable.

Regarding the within group comparisons, it should be noted that the dual goal and control task subjects displayed a pattern of prediction question performance identical to their equivalent sets of subjects from Experiment 1. However, the pattern search subjects display a pattern of data that is markedly different from the pattern search subjects from Experiment 1. The pattern search subjects from Experiment 1 performed comparably on all three prediction question types. The fact that only the pattern search subjects have a non matching data pattern could be explained by the fact that the data was collapsed over verbalisation. It was only for the pattern search subjects that there was a verbalisation condition (the 'describe' condition) that had been predicted to interfere with the normal pattern of data. Therefore it is likely that the odd pattern of data was due to the 'describe' pattern search subjects performing differently from the normal pattern of data and therefore when collapsed with the 'explain' pattern search subjects, it altered the pattern normally expected for pattern search subjects.

Inspection of Figure 3.3 (looking at the bottom chart) adds weight to this explanation. As expected, the 'explain' pattern search subjects appear to perform comparably on all prediction question types. It is for the incompatible 'describe' condition that there appears to be a difference in performance from the expected pattern. Subjects appear to be performing particularly poorly on New questions. Post hoc within group comparisons confirmed these observations. The 'explain' pattern search subjects did perform comparably on all question types whereas the 'describe' pattern search subjects did not. They performed worse on the New questions with performance on the Old-wrong question being significantly better, $Z = -2.24$, $p < 0.025$ (performance between Old-correct and New question and between Old-wrong and Old-correct questions was comparable).

As in Experiment 1 the results were also analysed defining the prediction situations by the last two elements (again discarding questions for which the situations could have occurred with both a correct and an incorrect response) - see page 31 for a more detailed explanation for why this was examined. The results can be found in Appendix 3 on pg. 239. These reanalyses showed an identical pattern of statistics with all the results that were significant before remaining significant (all p values < 0.05).

Correlations between control performance and predictions: To examine how predictions relied upon control performance, total prediction scores were correlated with the number of correct test trials. Spearman rank correlation coefficients were 0.87 for the 'explain' pattern search group ($p < 0.001$), 0.38 for the 'explain' control task group ($p < 0.3$), 0.50 for the 'explain' dual goal group ($p < 0.1$), 0.76 for the 'describe' pattern search group ($p < 0.005$), -0.18 for the 'describe' control task group ($p < 0.6$), 0.47 for the 'describe' dual goal group ($p < 0.2$). These results show the control task subjects showed no correlation between test phase performance and prediction performance. The two pattern search groups show a significant correlation between test phase and prediction performances. One of the two dual goal groups shows a high correlation between test phase and prediction performances however, it just fails to reach significance. For the groups with a specific, control goal the Spearman Rank correlations between total questionnaire performance and overall *learning* phase score were also examined. For the 'describe' control task group had a correlation of -0.61 which is significant ($p < 0.04$). For the 'describe' dual goal group had a correlation of 0.31 ($p < 0.4$). The 'explain' control task group had a correlation of 0.17 ($p < 0.6$). The 'explain' dual goal group had a correlation of 0.42 ($p < 0.2$).

The Rule Description Questions

The answers were judged by two judges and ranked into three categories; *No information or Wrong, Partially Correct, Correct.* The ranking procedure was identical to that used in Experiment 1 (see pg. 32). Both judges ranked the answers identically. These rankings can be seen in Table 3.1.

Table 3.1
Ranking of answers to the rule description question section into each category for the two groups.

Group: Goal	Group: Verbalisation	No information or Wrong	Partially Correct	Correct
pattern search	describe	**9**	**1**	**2**
	explain	**0**	**0**	**12**
control task	describe	**9**	**3**	**0**
	explain	**12**	**0**	**0**
dual goal	describe	**11**	**1**	**0**
	explain	**9**	**1**	**2**

Fisher exact probability tests were performed using the number of answers in the *No information or wrong* category and in the *Correct* category. The 'explain' pattern search group outperformed the other groups, getting significantly more answers in the *Correct* category and less in the *No information or wrong* category than every other group (every p value < 0.00001). Comparisons between every other group showed no significant differences, indicating that they all had a similar amount of answers in the two categories. These results indicate that the 'explain' pattern search group was better than every other group at producing answers that contained declarative knowledge and that the other groups performed at a comparable lower level. As suggested in Experiment 1, it might be argued that too strict a criterion was used to place answers in the *Correct* category and that with a looser criterion all the groups may have been more similar. With this possibility in mind Fisher exact probability tests were carried out, but this time the number of questions in the *Partially Correct* and C*orrect* categories were added together. Again Fisher exact probability tests showed an identical pattern to before.

Verbalising vs Not Verbalising: Comparisons With The Data From Experiment 1

To get a further insight as to the effect of the two different forms of verbalisation, comparisons were made with the non-verbalising groups from Experiment 1. For each goal group, scores from the two different sorts of concurrently verbalising groups were compared to the silent group from Experiment 1. Comparisons were made for the total learning phase score (where appropriate), test phase score and prediction score and also for the scores for the different types of prediction question. The F and *p* values of these comparisons can be seen in Table 3.2. For any non significant values, $p > 0.1$.

Table 3.2
Comparisons with the non-verbalising groups from Experiment 1

	Whole phase scores		Prediction questions			
	Learning	Test	All	Old-wrong	Old-correct	New
Exp. 1's pattern search group vs :						
'describe' pattern search group	NA	12.8 .001 ↓	6.55 .015 ↓	1.86 ns	3.66 .064 ↓	9.7 .004 ↓
'explain' pattern search group	NA	7.87 .008 ↑	5.02 .032 ↑	6.62 .015 ↑	3.38 .075 ↑	3.06 .089 ↑
Exp. 1's control task group vs:						
'describe' control task group	11.9 .002 ↓	2.55 ns	1.09 ns	0.03 ns	2.81 ns	0.009 ns
'explain' control task group	6.42 .016 ↓	0.7 ns	6.89 .013 ↑	3.86 .058 ↑	6.53 .015 ↑	0.17 ns
Exp. 1's dual goal group vs:						
'describe' dual goal group	5.11 .03 ↑	2.15 ns	3.79 .06 ↑	0.77 ns	2.77 ns	2.3 ns
'explain' dual goal group	1.9 ns	6.61 .015 ↑	8.32 .007 ↑	3.27 .079 ↑	6.85 .013 ↑	1.44 ns

'↑' indicates verbalising group outperforming Experiment 1's silent group, '↓' indicates the opposite. **F(1,34)** ***p* value** **ns = not significant** **NA = Not Applicable**

For the groups with a pattern search goal, the subjects who were concurrently explaining what they were doing performed better on all comparisons for the total and Old-wrong prediction question scores than the silent subjects from Experiment 1 and just failed to perform significantly better for the Old-correct and New scores. The silent subjects from Experiment 1 performed better than the subjects who were concurrently describing what they were doing, overall for the test phase and prediction questions and for the New questions and just failed to outperform them on the Old-correct questions. There was no significant difference in performance for the Old-wrong questions.

For the groups with a control task goal, the subjects who were concurrently explaining what they were doing performed better than Experiment 1's silent subjects on the Old-correct questions and just failed to significantly perform better on the Old-wrong questions and total prediction score. There was no significant difference in performance for the New questions or the total test phase score. For the total learning phase score the silent subjects outperformed the concurrently 'explaining' subjects. The silent subjects also outperformed the concurrently 'describing' subjects for the total learning phase score. For all the other comparisons, the concurrently 'describing' subjects performed comparably with the silent subjects.

For the groups with a dual goal, the subjects who were concurrently explaining what they were doing performed better than Experiment 1's silent subjects overall on the test phase, total prediction question scores and on the Old-correct questions and just failed to perform significantly better on the Old-wrong questions. Performance for the learning phase and on the New questions was comparable to the silent subjects. The concurrently 'describing' subjects performed significantly better than the silent subjects overall for the learning phase and just failed to perform significantly better for the overall score on the prediction questions. For the other comparisons the concurrent 'describing' subjects performed comparably with the silent subjects.

DISCUSSION

The aim of the study reported in this chapter is two fold. The objectives have been (1) to demonstrate that different learning goals result in different learning modes due to a direct influence of the goals on cognitive processes and, (2) demonstrate the positive effect of explanations on learning. For point (1), specifically the aim has been to demonstrate that a control task learning goal leads to instance learning because it encourages and confines subjects to the exploration of instance space, and a pattern search goal leads to explicit rule learning because it encourages subjects to explore rule space. The results support these propositions on how goals influence learning and the results also demonstrate the positive effects of explanations on learning. The concurrent secondary task designed to confine subjects to the search of instance space (the 'describing' task), did not impair control task subjects' performance - instance learning still occurred. However, for pattern search subjects, explicit learning of the underlying rule was prevented. A concurrent secondary task designed to encourage the exploration of rule space (the 'explaining' task), led to pattern search subjects' performance not only being unimpaired, but being enhanced. This demonstrates both the positive effect of explanations on learning and supports the proposition that the induced cognitive effect of a non-specific, pattern search goal results in a search of hypothesis space. Also, as predicted, for control task subjects, the 'explaining' secondary task did not alter the nature of subjects learning (they still learnt instances), but it enhanced their performance, again demonstrating the positive effect of explanations on learning.

For the rest of this Discussion, firstly, it is detailed exactly how the results of each goal group support the above conclusions. Then, other notable aspects of the results including those of the dual goal group are discussed. Finally the findings are discussed in relation to the explanation learning literature and general models of learning.

Subjects with a pattern search goal and an 'explain' secondary task clearly learnt explicit rules like their silent counterparts from Experiment 1. This can be concluded from (a) all subjects in the 'explain' pattern search group were able to describe the pattern that Clegg followed, (b) the subjects made predictions equally well, irrespective of whether they were predicting from experienced situations or from novel situations, (c) there was a significant positive correlation between prediction question score and test phase score. It is also clear that these subjects not only learnt rules as did their silent counterparts, but their performance excelled. This can be concluded from the fact that the subjects outperformed their silent counterparts on every 'total score' measure, both of control performance and prediction question performance and for the Old-wrong prediction questions and marginally outperformed their silent counterparts for the Old-correct and New questions. The lack of impaired performance by the 'rule space focusing' concurrent task supports the notion that a non-specific pattern search learning goal induces rule learning by encouraging the exploration of rule space. The enhancement in performance of the 'explain' secondary task further supports the notion that explanations enhance learning.

Subjects with a pattern search goal and a 'describe' secondary task did not explicitly learn the underlying rule that Clegg followed. This can be concluded from (a) 10 out of 12 subjects could not correctly state the rule, (b) comparisons between the prediction question types suggested that subjects found it hardest to make predictions from novel situations - their performance on the New questions was significantly lower than for Old-wrong questions. It is also clear that the secondary task impaired the normal performance of subjects with a pattern search goal. Apart from point (b) showing that their pattern of data was different from that normally expected for pattern search subjects, this is concluded as the 'describe' pattern search goal subjects significantly underperformed their silent counterparts from Experiment 1 overall on the test phase and prediction questions and also for the New questions. Also, they just failed to significantly underperform the silent subjects on the Old-correct scores. The impairment of performance of the 'instance space focusing' secondary task and its prevention of explicit learning supports the notion that a pattern search goal normally causes the learning it does because the goal encourages subjects to focus primarily on rule space.

Also as predicted, within group comparisons between the 'describe' and 'explain' pattern search subjects showed that the explaining subjects were better than the describing subjects overall during both the test phase and the prediction questions. This adds further support to the notion that a pattern search goal leads to the exploration of rules space. For, explaining is compatible with the exploration of rule space and describing is not. This difference in performance also reinforces the notion that self-explanations enhance learning.

Subjects with a control task goal and an 'explain' secondary task still learned instances in the same manner as did their silent counterparts. This can be concluded from (a) none of the subjects were able to state the rule Clegg followed, (b) subjects were able to make predictions from situations they had experienced before (equally well from Old-wrong and Old-correct situations), but *not* from novel situations. These subjects learn better than if they had no secondary task. This can be concluded from the fact that subjects outperform their silent counterparts from Experiment 1 on the Old-correct questions and on the total prediction scores and just fail to significantly outperform them on the Old-wrong questions. This improvement in learning caused by the 'explain' secondary task supports the notion that explanations enhance learning. The fact that the 'explain' secondary task did not lead to explicit rule learning supports the notion that a control goal induces instance learning as the goal confines subjects to the search of instance space.

Subjects with a control task goal and a 'describe' secondary task did not have their normal performance impaired. This can be concluded from the fact that (a) their pattern of performance on the prediction questions was identical to that of their silent counterparts, and (b) when compared to their silent counterparts there was no drop in performance on any comparison excepting for the overall score during the learning phase. This particular drop in performance was also shown for the 'explain' control task group, that otherwise outperformed their silent counterparts. It is probable that this drop in performance occurred during the learning phase as this was the phase that subjects had to start verbalising in. As it occurred in both verbalising groups, the drop therefore, was probably caused by the act of verbalising and was not related to what subjects had to verbalise. The lack of difference with the silent subjects in performance in the test phase was probably because verbalising subjects were comfortable with verbalising by this stage in the experiment. For the 'describe' control task group, the general lack of impairment of performance of the 'instance space focusing' secondary task supports the notion that a control goal causes the learning it does as subjects are encouraged to focus on instance space.

The dual goal groups with both the 'describe' and 'explain' secondary tasks performed better than their silent counterparts. This can be concluded as (a) for every comparison made there was no drop in performance compared to the silent counterparts, (b) subjects with the 'explain' secondary task actually outperformed the silent subjects on the overall test phase score and questionnaire score, and on the Old-correct scores and just failed to significantly outperform them on the Old-wrong scores, (c) subjects with the 'describe' secondary task outperformed the silent subjects on the learning phase score and just failed to significantly outperform them on the total prediction question score. As suggested in the introduction to this experiment, this improvement in performance of dual goal group subjects with a secondary task probably occurs because subjects now focus more on one of the dual goals. This extra focus on one goal allows enough resources to be focused upon that goal to go some way to achieving it. Therefore the results here support the notion that silent subjects with a dual goal perform the way they do because the dual goal puts too much demand on cognitive resources not allowing any explicit processes to occur. The fact that the 'explain' dual goal group outperforms their silent counterparts on more measures than the 'describe' dual goal group also supports the notion that explanations enhance learning.

It has been shown that the 'describe' pattern search group and the 'explain' dual goal group did not explicitly learn the underlying rule operating Clegg. However these groups did have a significant (or for the 'explain' dual goal group, marginally significant) positive correlation between test phase score and prediction score. As a point of methodological interest, this highlights the fact that although this significant positive correlation is expected for explicit rule learners, to provide evidence of explicit rule learning, more than this positive correlation is required. It must be seen in conjunction with other factors to clearly indicate explicit rule learning.

It has been clearly established that the 'explain' pattern search subjects learnt explicit rules like their silent counterparts. The exact nature of the learning of the remaining groups is now considered. For both the dual goal and control task subjects their normal patterns of learning as seen in Experiment 1 appear to have been retained. The within group comparisons for the prediction questions demonstrated this. As in Experiment 1, the dual goal subjects were better on Old-correct questions than either the Old-wrong questions or the New questions, which were performed comparably. Also, as in Experiment 1, the control task subjects performed comparably on both Old-correct and Old-wrong questions and were better on these than on New questions. For the control task subjects this is as predicted as both control task groups should still have been learning instances as neither of the verbalisation conditions were expected to alter the nature of subjects' learning.

For the dual goal subjects no such predictions were made. It was hypothesised that both verbalisation conditions would allow one of the dual goals to be sufficiently focused upon to lead to better performance results than those shown for the Experiment 1's silent subjects. As discussed above this was shown. However with their pattern of prediction questions remaining the same as the dual goal subjects from Experiment 1, it suggests that as for the control task subjects the effect of the verbalisation conditions were not strong enough to change the nature of their learning - they still learnt instances. It was suggested in Experiment 1 (and is investigated and confirmed in Experiment 4) that the dual goal learning goal condition leads to implicit learning. It is open to debate as to whether pure implicit learning is still occurring for the two verbalising dual goal groups. It is possible that they are learning through an unknown mixture of implicit and explicit processes as is suggested for control task subjects, because it is likely that any improvement in performance for the verbalising dual goal subjects is occurring through explicit processes. The verbalising dual goal subjects are supposed to be outperforming their silent counterparts due to the verbalisations leading to focusing more on one goal only and therefore reducing cognitive load by not having to deal evenly with two goals. Any performance increase therefore, relating to the reduction of cognitive load could not come from implicit processes as by definition they are unaffected by cognitive load. Unfortunately the exact nature of the learning of the verbalising dual goal subjects is uncertain.

This leaves the question of how the 'describe' pattern search subjects are learning. It has already been shown that they are not explicitly learning rules. It is likely that they are simply performing instance learning. There performance was worse on the New questions, typical of an instance learners' pattern. It can also be suggested that their instance learning is more similar to that of the control task subjects than the dual goal subjects. Their pattern of prediction question performance leads to this conclusion. Their similarities to the control task subjects are that they perform comparably on Old-correct and Old-wrong instances and perform better on Old-wrong instances than on New instances. It is also these two points that distinguish them from the dual goal instance learners.

A brief examination of the subjects' verbalisations revealed some predictable patterns of content. Some transcribed examples of the content of the verbalisations from the learning phase for the different groups can be seen in Appendix 8 starting at page 263. Examination of the 'describe' recordings revealed an identical pattern irrespective of learning goal. Subjects (as instructed) religiously gave a running commentary on what was occurring and never hypothesised about why events were occurring. The main area of interest therefore is how the learning goals affected subjects trying to explain what was happening. As predicted, the pattern search subjects were inclined to form hypotheses about the general underlying structure of the tasks, and reviewed their hypotheses during the learning phase. The control task subjects tended to form hypotheses about how to achieve their specific goal, not hypotheses about the general underlying task. The dual goal group subjects' verbalisations were more in line with the pattern search subjects, however they were clearly hampered in exploring their general hypotheses by constantly having to achieve a specific goal. These patterns are expected considering the likely influence the learning goals should be having on subjects' behaviour, and support the notions of how the different learning goals influence learning.

Before turning to the implications of the results for other learning models, the results and their considerations are briefly summarised. As predicted, the results showed that explanations facilitated the rule learning of the pattern search goal subjects, while either having a control task or dual goal with either secondary task, or giving descriptions, fostered instance learning.

It has been suggested that the results of the 'explaining' pattern search subjects occur as learners use a combination of empirical learning and rule learning, since, in keeping with the dual space model, the rule learning involves creation of hypotheses in rule space and then testing of these hypotheses in instance space. In the concept learning literature, Wisniewski and Medin (1995) have proposed a model in which empirical learning and theory driven learning interact. Machine learning researchers have also developed systems that combine both empirical and explanation based learning (e.g. Lebowitz, 1986).

However, the results reported in this experiment pose a problem for the concept learning models: how to explain the influence of learning goal or secondary task on the acquisition of instances on the one hand and rules on the other. The strongest evidence for this dissociation between instances and rules comes from the prediction questions. Only the 'explaining' pattern search subjects made correct predictions in both old and new situations, consistent with performance based on a rule. The remaining five groups, consistent with the retrieval of stored instances, gave more correct predictions in the Old-correct situations (and for the control task subjects gave more correct predictions also in the Old-wrong situations) than in the New situations. According to Wisniewski and Medin's interactive model, people will learn instances only when they have no prior knowledge to inform learning. However, in this study, it can be assumed that all subjects had roughly the same prior knowledge available to them. In line with Wisniewski and Medin's interactive model 'explaining' pattern search subjects presumably used their prior knowledge of mathematics to help them form, test and refine hypotheses. But, in disagreement with Wisniewski and Medin's interactive model, prior knowledge was not used by the 'describing' pattern search subjects or by subjects who had a control task or dual goal. The dual space models of Klahr and Dunbar (1988) and Simon and Lea (1974) give the best account of this observation, since in these models, learning can be directed to one or both problem spaces as a function of learning goal and type of verbalisation. In the absence of such direction, it is likely that relevant prior knowledge guides the learner to use the hypothesis space as well as the instance space, as was observed by Wisniewski and Medin (1995).

The dramatic improvement in the pattern search subjects who gave explanations testifies to the powerful effects of explanations on learning (e.g. Chi, et al, 1989; VanLehn, & Jones, 1993). The non-specific goal subjects who gave explanations in the present study learned considerably better than Experiment 1's silent subjects. For example, 100% of the 'explaining' pattern search subjects gave correct rule descriptions, while only 76% of Experiment 1's silent subjects gave either complete or partial descriptions. In the educational literature, Ng and Bereiter (1995) have identified three kinds of learners who each spontaneously adopt a different learning goal. Learners with performance goals focus on completing the learning tasks. Such learners can be equated with what Stevenson and Palmer (1994) call 'learning through problem solving'. Learners with instructional goals focus on the manifest learning objectives; they use their background knowledge to help them understand the material but do not use the new material to restructure prior knowledge. This kind of learning can be equated with what Stevenson and Palmer call 'learning through memorisation'. Finally, learners with knowledge building goals focus on going beyond the instructional material in pursuit of wider learning goals. Only these learners use the new material to restructure prior knowledge as well as using prior knowledge to understand the new material. This kind of learning can be equated with what Stevenson and Palmer call 'learning through understanding'.

While these three kinds of learning are not mutually exclusive, it may be speculated that 'describing' control task subjects were learning through problem solving; they searched instance space for a route to the goal. It may also be speculated that the silent pattern search subjects from Experiment 1 were learning through memorization. They used prior knowledge in conjunction with the initial learning instances to construct a possible hypothesis but may have done little revision of the hypothesis in the light of subsequent learning trials. Finally, the 'explaining' pattern search subjects in the present study seem to have been learning through understanding. Giving explanations seems to have encouraged them to modify and refine their hypotheses until the underlying rule was correctly acquired. The findings, therefore, suggest ways in which learners can be guided to learn more effectively, since goal orientation and the use of explanations can be modified to the advantage of the learner.

The power of explanations was also demonstrated for the control task and dual goal subjects. As mentioned above, the fact that the 'explain' dual goal subjects outperformed their silent counterparts on more variables than the 'describe' subjects again demonstrates the positive effect on learning of explanations. It is possible to speculate that the dual goal explainers' performance improved through explanations due to similar mechanisms as that attributed to the pattern search explainers. That is, due to the constant reappraisal of hypotheses. For, if the dual goal explainers are essentially focusing on the goal compatible with their verbalisation condition then this would make their circumstances most similar to the pattern search explainers. The control task explainers are probably benefiting from the effect of explanations for different reasons. They are not pursuing a rule detecting goal. It was suggested that they may benefit from explanations as it may encourage them to more rigorously adopt means ends analysis to achieve the specific goal. Exactly how explanations enhance the learning of the control task and dual goal subjects is clearly open to debate. The important point is that it has been demonstrated in this study for instance learners, something that has not been specifically examined or demonstrated before.

In summary, the results indicate that learning goals *directly* influence learning by encouraging learners to explore instance space alone, or rule space and instance space. The Discussion examined the results of each group and demonstrated how these results support the conclusions. The study also explored the effects of explanations on the different forms of learning induced by the different learning goals. As predicted self-explanations improve learning for explicit rule learners. They also appear to have some positive effect for instance learners. Both the control task and dual goal groups outperformed their silent counterparts on some measures. With regards to the dual goal group subjects the results show that with either secondary task, performance improved compared to their silent counterparts. This was taken to support the notion that dual goal group subjects, when silent, have a heavy cognitive load that prevents them from performing any explicit learning. The effect of having a secondary task directed subjects to either one of the goals and so allowed enough resources so that some headway could be made with the particular goal. Finally the results were discussed in terms of other models of learning such as Wisniewski and Medin's (1995) interactive model of concept learning. It was concluded the dual space model of learning is the best model to explain all the results.

Chapter IV

Experiment 4: Implicit Learning

Experiment 1 provided evidence for three different types of learning from three different learning goals. Explicit rule learning was shown by subjects given the non-control oriented, pattern search goal. The specific[3] and dual goal conditions led to two different sorts of instance learning, the former where both correct and incorrect instances are memorised and the latter where only correct instances are memorised. It was proposed that this latter form of instance learning is occurring purely implicitly. It was reasoned that the dual goal subjects were torn between two different concurrent and conflicting cognitive processes. In terms of the dual space model - one required the search of instance space alone and the other required the search of hypothesis space as well. It was suggested that these conflicting goals led to a working memory overload preventing subjects from learning explicitly. The study presented here investigates this suggestion.

In this study, subjects are given the same 'Clegg' version of the person interaction task, but this time, to overload the central executive, subjects have a secondary task of concurrent random number generation. Below, the comments in chapter 1 that supported the idea of the dual goal group learning instances purely implicitly are summarised and enlarged upon. Following this the relevant work on dual tasks and working memory is addressed. Finally the new study is introduced and the hypotheses presented.

The notion that subjects in the dual goal group are learning instances by only memorising trials they perform correctly has support from the work of Dienes and Fahey (1995). Dienes and Fahey reworked a paper by Marescaux, Luc, and Karnas (1989) that explored the possibility of using a look-up table model to the describe instance learning of a dynamic system (the same as that used throughout this thesis). Dienes and Fahey improved on Marescaux et al's study by comparing results to a baseline, including the equivalent of New prediction questions in a post learning task measure, and also, by explicitly testing the suitability of a look-up table to model the results. Dienes and Fahey's experiment set subjects a version of the dynamic system used in this thesis (the sugar production task - as used in Experiments 6a and 6b). Following the task subjects were given what was called the specific situations task. This

[3] Also referred to as the control task goal group in Experiments 2 & 3.

was similar to the prediction questions used in this thesis in that subjects were presented with some trials that were similar to the ones they had previously experienced. However, rather than asking the equivalent of 'where would you expect Clegg's next response to be?', they were asked, 'what input would you enter to make Clegg produce the required output of the testing phase?' The results of subjects performing on Old specific situations were measured in terms of concordance. Concordance was defined as the percentage of times subjects entered their old response they had previously used in that situation. The learning was modelled by saying subjects built a look-up table from the instances they encountered and recorded Old-correct instances in this table. The conclusion that subjects only entered Old-correct instances into the look-up table came from the fact that subjects showed a higher concordance for Old-correct specific situation task questions than for Old-wrong ones.

So, Instance learning can be described by subjects building a look-up table. Some of the instances subjects encounter are entered into the look-up table (e.g. Logan, 1988). As to which sorts of instances (i.e. instances of correct or incorrect responses) are entered into the look-up table depends on the theory being described. As described above, Dienes and Fahey tested a look-up table model based on Logan's instance theory and found the best fit with the data occurred when the look-up table stored correct instances only. This matches the dual goal group's data as it would be the way to construct a look-up table to produce its learning pattern. The other obvious option is to make entries in the look-up table of all instances encountered (i.e. from both correct and incorrect performances). This sort of look-up table could be used to describe the specific goal group's pattern of learning.

So, two different forms of look-up tables can be used to describe the two different forms of instance learning resulting from dual and specific goals. Examination of the mechanics of these two different forms of look-up table highlight processing and theoretical reasons that add support to the notion that the dual goal group is learning implicitly. Firstly, consider a look-up table model that describes the dual goal group's learning. The look-up table's store of correct instances can be referred to for actions that achieve the goal. For the use of correct instances, a subject simply identifies a match between an instance in the look up table and the present position in the task and then repeats the prescribed action. In other words, the look-up table is used to drive the subjects' actions by simply repeating previous successful actions. Notably, there is a *direct* link between (a) the recognition of the match between instance in the look-up table and task situation, and (b) action needed to produce the required goal. Now, consider a look-up table used to produce the explicit instance learning of the specific goal group. This has an additional store of incorrect instances. The store of incorrect instances can be referred to for actions known *not* to achieve the goal. However, to use incorrect instances to guide future actions, extra steps are needed: As in the use of correct instances, the subject identifies a match between an instance in the look up table and the present position in the task. However, the subject must then make the additional step of realising that the previous actions did not lead to correct performance. Then, the subject must decide upon a different action from that used before and finally execute that action. These extra steps require processing power that does more than simply tie an instance match with a prescribed action. It is predictable therefore, that with a cognitive load, the processing cost is more likely to interfere with incorrect than with correct instances.

If the cognitive load gets too heavy, one might expect the use of correct instances to be interfered with as well. However, the use of correct instances uses a direct link between match of instance and the prescribed action. It is this direct link between instance match and action that is at the core of why the use of correct instances may be implicit and hence unaffected by a heavy cognitive load. This direct link with a prescribed action can be seen as being purely procedural. According to Anderson's ACT model of learning (Anderson, 1983), tasks that are proceduralised no longer rely on components of working memory. In other words, there is reason to suggest that the direct link between instance match and action is automatic and therefore an additional cognitive load should not interfere in its function. The study presented here tests the idea that the learning of correct instances is automatic/implicit.

Next, relevant work on the dual task paradigm is discussed. Concurrent tasks have been used to examine working memory and implicit learning with a number of different paradigms used to explore implicit learning. For a detailed overview of the literature, see Seger, 1994, pg.177. Specifically, for sequence learning see: Cohen, Ivry & Keele, 1990; Keele & Jennings, 1992; Nissen & Bullemer, 1987. For artificial grammar learning see: Dienes, Broadbent & Berry, 1991. For control of dynamic systems see: Hayes & Broadbent, 1988; Green & Shanks, 1993; Sanderson, 1990, Porter, 1986, 1988. (Some of the pertinent work on dynamic systems, being most relevant to this study, is expanded on below.) The working memory model that is almost universally adopted in these and other dual task studies is that of Baddeley and his colleagues.

Baddeley's working memory model is perhaps the most widely used and tested model of working memory in cognitive science (Baddeley, 1986, 1990, 1992; Baddeley & Hitch, 1974; Logie, 1991; Logie, Zucco & Baddeley 1991). It consists of three main components, the visuospatial scratch pad, the articulatory loop and the central executive. The former two components are labelled as slave systems to the central executive. The central executive, also referred to as the attentional controller, is the component of most importance to the study presented here. It is the component at the heart of working memory and therefore controller of explicit learning processes. For the purpose of this study a secondary task was needed that would be dependent on the central executive and therefore interfere with any explicit learning processes in the primary task. The task used for this is random generation (see e.g. Gilhooly, Logie, Wetherick & Wynn, 1993; Hayes & Broadbent, 1988; Green & Shanks, 1993; For a detailed account of why random generation is deemed to occupy the central executive see Baddeley, 1990).

In the literature on dynamic systems, Hayes and Broadbent's study is worthy of particular attention because, like the study presented here, it uses both a version of the person interaction task and concurrent random number generation tasks. Hayes and Broadbent manipulated subjects' learning to make it supposedly implicit or explicit by adjusting the salience of what subjects had to learn. Both sets of subjects performed the person interaction task. However, the designated implicit learners had a more complex underlying equation operating the computer person than the designated explicit learners. After an initial learning period (adjusted in length so that the subjects with the tricky equation had longer to reach an equal level of performance) the underlying equation of both groups of learners was adjusted in an identical manner. The subjects with the simpler equation were better able to cope with this change than subjects with a complex equation. However, if all subjects had to randomly generate letters (or numbers in experiment 3) concurrent with the trials following the equation change, the opposite occurred and the supposed implicit learning group (subjects with the complex equation) were better able to cope with the change. It was reasoned that the secondary task overloaded working memory preventing explicit learning and therefore the designated explicit learners were no longer able to cope with the task change. However, due to the explicit functioning being stopped, this allowed implicit learning to function better and therefore cope with the task change. The study was concluded to demonstrate the existence of dissociable learning systems. Additionally, the study was thought to highlight how salience could be used to adjust subjects' mode of learning.

Unfortunately, there was a partial failure to replicate Hayes and Broadbent's results by Sanderson (1990), and a total failure by Green and Shanks (1993). Apart from failing to replicate the results, Green and Shanks provided evidence that the two tasks in Hayes and Broadbent's paper "[do not] induce two different modes of learning. Rather, the findings support a different position: that the two tasks differ only according to level of difficulty (pg. 314)". Unfortunately, as commented in the discussion of chapter 1 (see pg. 39), a key problem in adjusting subjects learning by adjusting the salience of the task, is that apparent differences in modes of learning can always have this alternative explanation. That is, that differences in behaviour are due to differing degrees of difficulty of the different tasks used to induce, supposedly, different modes of learning.

It is clear then, that in the field of dynamic systems, there is a distinct lack of successful experiments using the dual task paradigm to demonstrate or even explore implicit learning. Indeed Berry (Berry and Dienes, 1991, pg. 35) notes that "Very few studies have been published to date looking at the effects of secondary tasks on control task performance....Clearly, more studies are needed on this question.". The study reported here will go some way to addressing this problem. This study will be similar to parts of Hayes and Broadbent's study. It is arguable that evidence for different, dissociable learning systems has already been provided in Experiment 1. The purpose of the study here is to clarify whether or not one of the induced forms of learning (that of the dual goal group) requires working memory. In other words, the question is being asked as to whether or not the learning is implicit. To do this, as in Experiment 1 and 3, there are 30 learning trials with the 3 different learning goals. Then there are the 30 test trails where control performance of the computer interaction task is measured. Following this there are the 15 prediction questions that can be used to indicate either that rule learning has occurred or that instance learning (either implicit or explicit) has occurred. Specifically, new as well as old situations and both correct and incorrect old situations should yield accurate predictions if rule learning has occurred. If explicit instance learning has occurred then both correct and incorrect old situations but not new situations, should yield accurate predictions. If implicit instance learning has occurred then only correct old situations should yield correct predictions. Finally subjects are given the rule description questions, which can be used to further indicate whether subjects have any explicit knowledge of the underlying rule. For this study, the key additional change is that during the learning trials all subjects have to randomly generate digits at the same time. Digits were chosen as opposed to letters so as not to conflict in any unintentional manner with the primary task (as subjects' input to the computer is from a choice of letters). To complete the design there are also another three groups of subjects performing the secondary task during the test trials as well as during the learning trials.

The hypotheses: In summary, the combination of a specific control goal and a pattern search goal for the dual goal subjects places a cognitive load that is too heavy for subjects to engage in explicit learning. Therefore, the pattern of data that the dual goal group displays is purely that of implicit learning processes. Logically, the extra cognitive load of a concurrent task should not impair the dual goal group subjects, and therefore they should display the same pattern of data. Furthermore, if subjects with just the specific control goal or subjects with just the pattern search goal were given a heavy additional cognitive load then they would only be able to engage in implicit learning processes. Consequently the control task and pattern search goal subjects should display the same pattern of data as the dual goal group.

More precisely it is expected that all the groups presented below will display the original dual goal group data pattern - that of pure implicit instance learning. Therefore, subjects should show no correlation between prediction questions and test trials performance. In answer to the rule description questions, they should show no evidence of verbalisable knowledge of the underlying pattern the system follows. Critically, they should excel at prediction questions where they are predicting from situations they have encountered before and performed correctly in. They should perform equally poorly when predicting from familiar incorrect and novel situations. As all groups have the additional cognitive load of the secondary task, they should perform comparably on all measures, irrespective of learning goal (control task, pattern search or dual).

METHOD

Subjects: The 72 volunteer subjects were Durham University graduate and undergraduate students, aged between 18 and 24.

Design: A 3 (goal) by 2 (locus of secondary task) independent groups design was used. The 3 goals were the specific control goal, the pattern search goal and the dual goal (both control and pattern search). The secondary task was performed either during the learning trials only or during both the learning and test trials. Subjects were randomly allocated to one of three goal groups (control task, pattern search and dual goal). All subjects were required to complete 30 learning and 30 test trials. Half the subjects in each goal group performed the secondary task during the learning trials only and the other half performed it during both learning and test trials. In the test trials, all subjects were given a new specific goal. There were two specific goals: to make Clegg *Polite* and to make Clegg *Very Friendly*. The specific goal for the learning trials was to make Clegg *Polite*. The specific goal for the test trials was to make Clegg *Very Friendly*. After the test trials, all subjects were given an unexpected questionnaire, consisting of 15 prediction questions followed by two rule description questions, which probed explicit knowledge of the rule. There were 3 types of prediction question - 5 based on old correct trials, 5 on old incorrect trials and 5 on new trials. These 15 questions were displayed in a random order. The order of the two rule description questions was counterbalanced across subjects .

The task : Subjects learned the Clegg version of the person interaction task which was identical to that described in chapter 1 (see pg. 21).

In addition to explaining the nature of the task, each group of subjects was given specific instructions concerning their learning goal. These were the same as those given to the different goal groups in Experiment 1. Also, all groups of subjects were given identical instructions concerning their additional concurrent secondary task. They were told that they should call digits (0-9) out loud in a random way while doing the person interaction task. To convey the meaning of random, subjects were told to imagine that the numbers 0,1,2,3,4,5,6,7,8,9 were written on separate pieces of paper, placed into a hat, mixed up and then a piece of paper was taken from the hat and called out. The piece of paper was then mixed back into the hat and a piece of paper was again drawn from the hat and so on. To illustrate the task the experimenter then called out some digits for a few seconds. They were deliberately not told to call out the numbers at any particular pace. This method was used as opposed to subjects calling out numbers to the rhythm of a metronome. Bourke, Duncan and Nimmo-Smith (1996) have recently noted from the results of a pilot study of theirs, that the metronome method can lead to subjects almost completely neglecting the other task, however, getting subjects to call numbers at their own free will prevents one task from dominating over the other. Subjects were also warned that they would be encouraged by the experimenter to call out digits if they stopped for long periods. Subjects were given a minute or so to practise generating random numbers.

The test trials for all subjects, were identical to the learning trials of the control task group except that the goal was changed. The subjects in each group had to make Clegg *Very Friendly* and maintain him at that level. As was the case in the learning trials, a response either on the target or one step either side of the target was scored as correct, to allow for the random element in the equation. Half the subjects in each goal group were told that they did not need to generate random numbers during the test trials, the remaining halves were told that they should continue to generate random numbers.

The Questionnaire : this consisted of prediction questions and rule description questions identical to those described in the Method in chapter 1 (see pg. 23).

Procedure: Subjects were randomly allocated to one of the three goal groups and then randomly allocated to one of the two secondary task conditions. As mentioned above, the three groups received identical initial instructions apart from one sentence. This sentence dictated the aim of that particular group for the person interaction task during the learning trials. Apart from this, the remainder of the experiment was identical for all groups:

The instructions explaining the nature and aim of the subjects' initial learning task were presented first. Additionally instructions explaining the concurrent secondary task were presented. Subjects had about a minute to practise generating random digits. This was followed by the learning trials. On completion of this phase, subjects received instructions describing their new goal for the test trials and then the test trials started. The groups were also given their respective instructions as to whether they should continue generating random numbers. Clegg initiated both learning and test trials by displaying one of the three adjectives centred on *Polite*. Following the test trials the subjects were presented with instructions for the prediction questions. These instructions described the nature of the questions and gave an example of a situation from which the subjects would have to make a prediction. The instructions also explained that each question was unrelated to the previous one. After completing the prediction questions subjects were given a pen and paper and were asked to answer the two rule description questions appearing on the paper.

RESULTS

Learning Trials

As mentioned before, learning trials were scored as correct for the control task and dual goal groups if they got a response from Clegg of *Indifferent, Polite* or *Very Polite*. This scoring takes into account the random element of the equation producing Clegg's behaviour. Due to the lack of a specific goal for the pattern search subjects during their learning phase, no measure could be made for their performance during the first set of trials. The mean number of correct learning trials for these groups are shown in Figure 4.1. In the Figure, data are shown for all 30 learning trials combined and for each half of the learning trials.

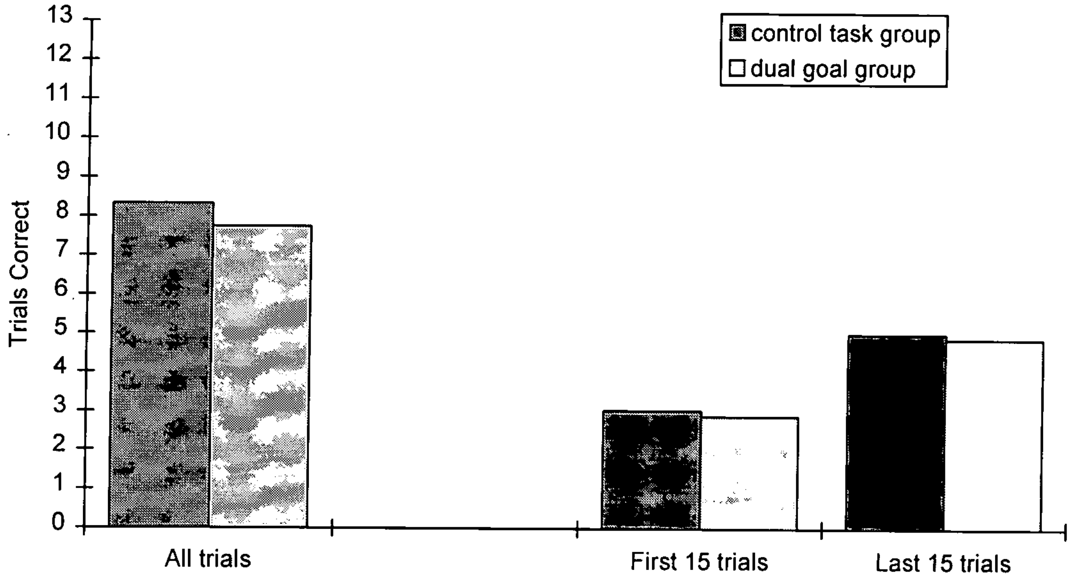


Figure 4.1: Mean number of correct trials in the learning phase for the control task and dual goal groups. Data are shown for all 30 trials combined and for the first and second 15 trials. Data were collapsed over secondary task as all groups concurrently verbalised during the learning phase.

The data in Figure 4.1 were analysed using a 2 (learning goal) by 2 (trial block: first 15 trials vs last 15 trials) analysis of variance with repeated measures on the last factor. There was a main effect of trial block, $F(1,44) = 13.68$, $p = 0.001$: subjects performed better in the second 15 trials than they did in the first 15. The results showed no main effect of learning goal. The interaction between learning goal and trial block was not significant. It is clear then that overall during the learning phase the two different learning goals did not affect performance. There was general improvement during the learning phase and this held true for both control task and dual goal subjects. (See Appendix 4 for the ANOVA tables and full sets of t-tests for this experiment, pg. 244).

Test Trials

For all groups, trials were scored as correct during the test phase if subjects got a response from Clegg of *Friendly, Very Friendly* or *Affectionate.* Figure 4.2 shows the mean number of correct test trials for each group for the entire test phase and each half of the test phase.

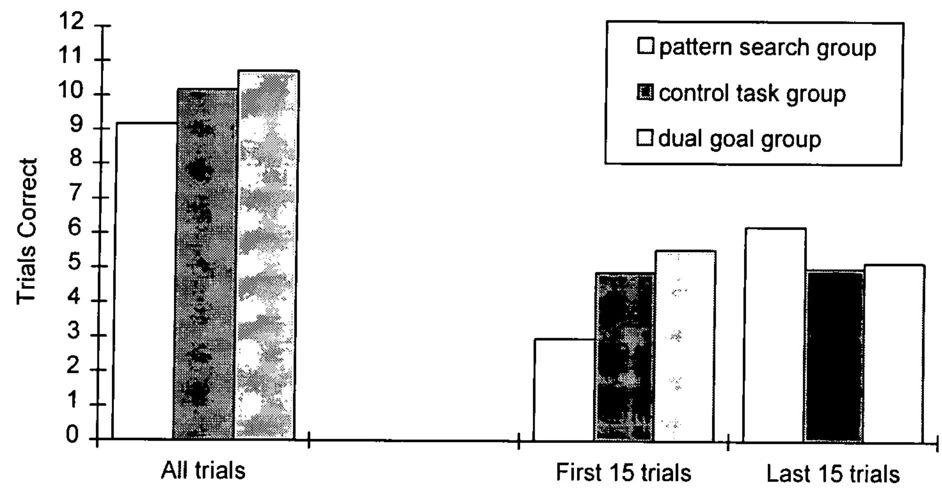


Figure 4.2: Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials. Data were collapsed over secondary task as it had no effect on performance.

The data in Figure 4.2 are analysed using a 3 (learning goal) by 2 (secondary task) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results showed no main effect of learning goal or secondary task. The interaction of the two main effects also was not significant. The main effect of trial block just failed to reach significance, $F(1,66) = 3.6$, $p = 0.062$. There was no significant interaction between trial block and secondary task, however there was between trial block and learning goal, $F(2,66) = 4.63$, $p = 0.013$. Consequently, with groups collapsed over secondary task, within groups comparisons were performed between the first and last half of the test phase. There was no significant difference between the first and last half of the test phase for the control task or dual goal groups, however, there was for the pattern search goal group t(22) = -3.55, *p* = 0.001. So, overall during the test phase there was no effect of learning goal. However the learning goal did effect subjects performance during the test phase. Only subjects with a pattern search goal showed an increase in performance during the test phase. Whether or not subjects had to perform a secondary task during the test phase did not affect their performance.

Transfer: In keeping with the rest of the thesis, another important issue to examine is how the control task and dual goal groups coped when the subjects switched goals from the learning phase to the test phase. Two comparisons were made to examine this issue: (i) A comparison between the number of correct trials during the last *half* of the learning phase and the first *half* of the test phase and, (ii) A comparison between the *total* number of correct trials during the learning and test phases. For comparison (i) a 2 (learning goal) by 2 (secondary task) by 2 (trial block: last half learning phase vs first half test phase) mixed analysis of variance revealed no main effects or significant interactions. This suggests that subjects' performance was not getting better or worse just after their specific goal changed. The lack of interactions suggests that this held true irrespective of learning goal or in how many phases subjects performed the secondary task. For comparison (ii) a 2 (learning goal) by 2 (secondary task) by 2 (phase score: learning vs test) mixed analysis of variance revealed a main effect of phase score, F(1,44) = 16.35, *p* < 0.001: subjects performed better in the test phase than in the learning phase. The three way interaction between secondary task, learning goal and phase score was marginally significant, F(1,44) = 3.22, *p* = 0.08: there was a tendency for subjects with a dual goal who were verbalising in the test phase to perform only comparably and not significantly improve between learning and test phases. There were no other significant effects or interactions. Overall the change in specific goal did not damage performance. Subjects were able to improve enough to perform significantly better in the test phase.

The Prediction Questions

These were scored in the same way as in Experiment 1 (see pg. 28). As in previous experiments, there is a potential hazard with the prediction questions that could undermine any conclusions based on their related statistics. Old-correct and Old-wrong questions were selected from the test trials, but any given situation could have occurred more than once and the response on another occasion might have been different from the one given in the selected situation. Therefore, a trial that was selected as an Old-wrong one might have been responded to correctly on another occasion, or vice versa. Such responses (9% of the data) therefore were discarded. The resulting mean percentage of correct responses are shown in Figure 4.3. Due to some of the data being discarded, these results are shown as percentages.

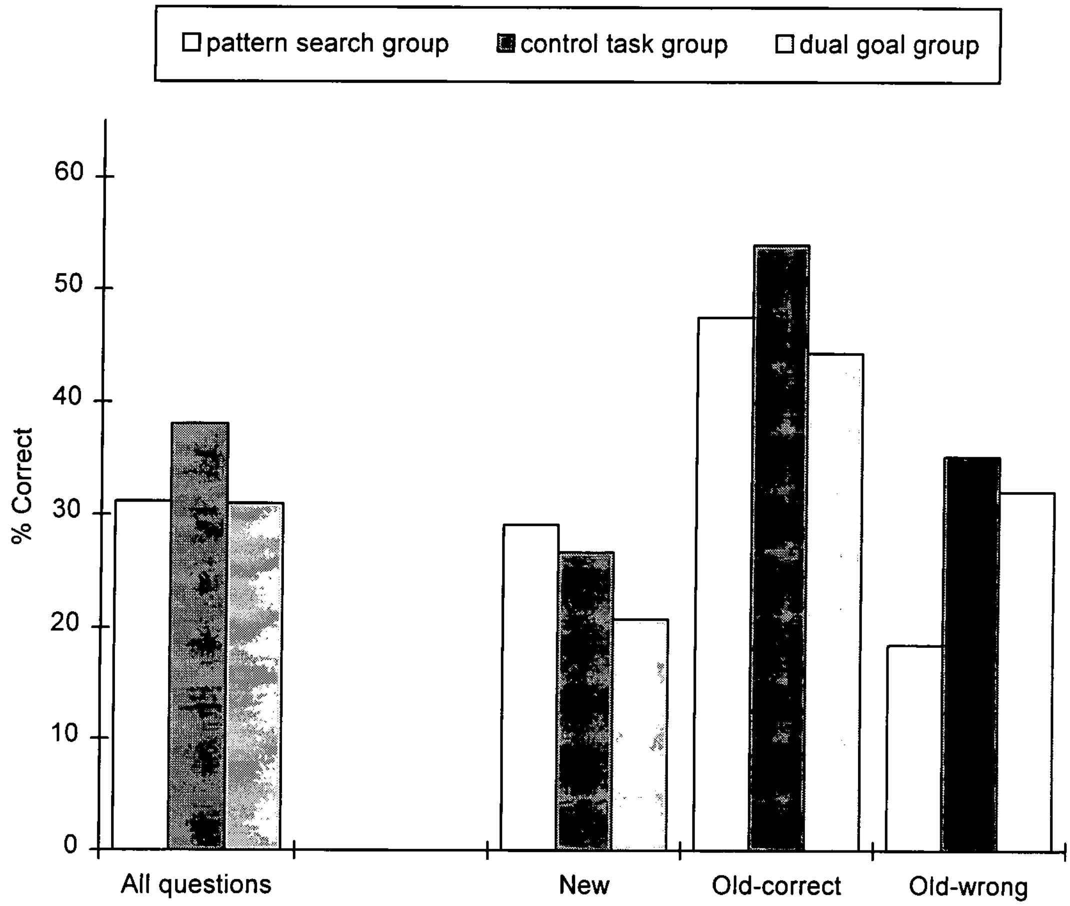


Figure 4.3: Mean percentage of correct responses to each category of prediction questions for each group. Data were collapsed over secondary task as it had no effect on performance.

The data in Figure 4.3 are analysed using a 3 (learning goal) by 2 (secondary task) by 3 (question type) analysis of variance with repeated measures on the last factor. The results revealed was a main effect of question type, $F(2,132) = 18.2$, $p < 0.001$. There was no other main effects and no significant interactions. With data collapsed over secondary task and learning goal, the average scores for the three question types were Old-correct, 48.63%, Old-wrong, 28.61%, and New, 25.56%. Wilcoxon matched paired tests showed that subjects performed significantly better on the Old-correct scores than on the New scores ($Z= -4.41$, $p < 0.0001$) and the Old-wrong scores ($Z= -3.83$, $p = 0.0001$). The comparison between the Old-wrong and the New scores failed to reach significance ($Z= -0.87$, $p = 0.38$). Overall, all groups performed comparably - different learning goals or carrying out the secondary task just in the learning phase or in both learning and test phases did not affect performance on the prediction questions.

As in Experiment 1, the results were reanalysed defining the prediction situations by the last two elements (again discarding questions for which the situations could have occurred with both a correct and an incorrect response) - for a longer explanation of why this extra analysis was done see pg. 31. The results of these reanalyses can be found in Appendix 4, pg. 245. These reanalyses showed an identical pattern of statistics with all the results that were significant before remaining significant (all p values < 0.05).

Correlations between control performance and predictions: To examine how much the predictions relied upon control performance, total prediction scores were correlated with the number of correct trials during the test phase. None of the correlations were significant in any of the 6 groups. This suggests that for all the groups performance during the test phase was not indicative of subsequent prediction performance.

The prediction scores for the control task and dual goal groups were also correlated with the number of correct trials during the *learning* phase. Again, none of the correlations were significant.

The Rule Description Questions

Subjects' answers to the two rule description questions were ranked in the same way as described in chapter 1 (see pg. 32). Both judges categorised the answers identically. These categorisations can be seen in Table 4.1.

Table 4.1
Numbers of subjects giving each category of response to the rule description questions

Goal	Locus of secondary task	No information or Wrong	Partially Correct	Correct
control task	learning phase	**11**	**1**	**0**
	learning & test phase	**11**	**0**	**1**
pattern search	learning phase	**12**	**0**	**0**
	learning & test phase	**11**	**1**	**0**
dual	learning phase	**10**	**2**	**0**
	learning and test phase	**11**	**1**	**0**

As can be seen from the data in Table 4.1, all groups had more answers in the *No information or wrong* category than the *Correct* category. Fisher exact probability tests comparing the number of answers in the *No information or wrong* category and in the *Correct* category showed that all the groups performed comparably with each other. It is clear then that none of the groups were producing answers that contained declarative knowledge about the underlying pattern of the person interaction task.

Random number generation

The measure of randomness used was Evans's (1976) RNG index (as used by Gilhooly et al, 1993) which gives a value between 0 and 1 where the lower the value the more random the series of digits. For the learning phase, data were collapsed over secondary task as all groups concurrently verbalised. The mean RNG values for the three goal groups were 0.27, 0.29 and 0.29 for the pattern search, control task and dual goal groups respectively. The mean number of digits produced during the learning phase for the three goal groups were, 144.13, 151.42 and 152.13 for the pattern search, control task and dual goal groups respectively. Two separate one-way analysis of variance for the RNG values and for the number of digits produced showed there was no difference between goal groups for either comparison, $F(2,71) < 0.6$. To examine whether there was any relationship between how random subjects were being or how many digits subjects were producing and their control and prediction question performance, Spearman Rank correlation coefficients were calculated examining how total learning and test phase scores and total prediction question scores correlated with both the number of digits generated and with the RNG values. None of these coefficients were significant.

The mean RNG values for the three goal groups that were generating random numbers during the test phase were 0.28, 0.28 and 0.30 for the pattern search, control task and dual goal groups respectively. The mean number of digits produced for the three groups were, 143.92, 146.83 and 175.75 for the pattern search, control task and dual goal groups respectively. A one-way analysis of variance for the RNG values showed there was no difference between goal groups, $F(2,35) = 1.1$, $p = 0.35$. A one-way analysis of variance for the number of digits produced also showed there was no difference between goal groups, $F(2,35) = 1.42$, $p = 0.26$. As for the learning phase, Spearman Rank correlation coefficients were calculated examining how total learning and test phase scores and total prediction question score correlated with both the number of digits generated and RNG values. None of these coefficients were significant.

All in all, the learning goal had no effect on subjects' ability to be random or on the quantity of digits they produced. Likewise, subjects ability at being random or the quantity of digits they produced had no effect on either their control ability or their prediction ability. In other words, how much they concentrated on their number generating task did not effect one way or the other there learning of the person interaction task. This is exactly what would be expected if subjects were learning implicitly.

Comparisons with the dual goal group in Experiment 1.

Finally, an examination was made of how the dual goal group in Experiment 1 compared with the groups reported in this study. Comparisons were made for the total scores in the learning and test phases and in the prediction questionnaire and for the different question types in the prediction questionnaire. For all these measures, analysis of variance from the groups in this experiment have shown no effect of learning goal or secondary task or an interaction between the two. Consequently, the data was collapsed over these 2 factors in order to compare the data with the data from the dual goal group of Experiment 1. Separate analysis of variance for the different measures showed no significant differences on any comparison. This indicates that the groups in this experiment performed comparably with the dual goal group of Experiment 1.

DISCUSSION

The aim of the study reported in this chapter is to demonstrate that subjects given a combined specific control goal and pattern search goal display pure implicit instance learning. The results show that the pattern of data from the dual goal group in Experiment 1 does indeed reflect pure implicit learning. In the present study, when subjects were given a concurrent task that occupied working memory they were still able to demonstrate the same pattern of data as that of the dual goal group of Experiment 1. This held true whether subjects' primary task was the specific control goal of the control task group, the pattern search goal of the pattern search group or the combination of these two goals in the dual goal group.

The conclusion that all the groups replicated the pattern of data from the dual goal group in Experiment 1 comes from the following points. (1) none of the subjects showed a significant correlation with prediction question score or learning or test phase scores. (2) None of the groups showed evidence of verbalisable knowledge on the rule description questions. (3) With the exception of within group performance during the test phase, subjects' different learning goals never had any differential effect on any of the results indicating that all groups were performing comparably at all stages of the experiment. (4) The groups excelled at prediction questions where they were predicting from situations they had encountered before and performed correctly. They performed equally poorly predicting from situations they had encountered before and performed incorrectly and in situations they had never encountered before. (5) When the data of all the secondary task groups of this study were compared to that of Experiment 1's dual goal group no significant differences were shown.

As mentioned above, the only difference between the three goal groups was the pattern search group's significant improvement during the test phase when the other groups didn't improve. It is most likely that this was due to the fact that unlike the other groups, the pattern search group had not had a specific goal up to this point in the experiment. It is likely that this took some time getting used to and therefore, this group and this group only demonstrated a marked improvement during the test phase. The other groups were used to the specific goal so had already reached a steady performance and the groups did not show a marked improvement. This is not to say that the other two groups did not continue to improve during the test phase as both groups performed significantly better during this phase than during the learning phase.

Of key importance to the methodology of the study reported in this chapter is that the secondary task sufficiently interfered with working memory to prevent explicit processes. It is important to have this firmly established so when the pattern of data reported here is scrutinised, one can be confident that the foot print of purely implicit learning processes is under examination. There is much evidence to conclude that the concurrent random generation task sufficiently interfered with working memory so as to completely prevent successful explicit learning processes: (a) None of the groups were able to demonstrate any level of explicit knowledge of the underlying equation of the task in answer to the rule description questions. (b) There was no significant positive correlation between test trials and overall prediction questionnaire score. Arguably, this is a vital and missing demonstrator of explicit knowledge. (c) As shown in Experiment 1, when the task is run without a concurrent task the different learning goals have a differential effect on control performance during both learning and test trials and also on prediction questionnaire performance. These effects were completely removed when the secondary task was introduced. (d) Finally, there was a total lack of difference at every point in the experiment between having the secondary task during just the learning phase and having it during both learning and test phases. This is indicative of the secondary task being strong enough to have already had its maximum impact during the learning phase. In other words, additional interference on working memory by the secondary task during the test phase was not possible as it had already completely blocked explicit processes. From the above points, it can therefore safely be concluded that the pattern of data shown in this study is of purely implicit learning processes.

There are some practical implications of the data shown here. There was no difference between having a secondary task and not for the dual goal group. Logically, this implies that the dual goal group was learning implicitly. It has been suggested that the conflicting goals of the dual goal group interfere with each other and prevent explicit learning taking place. It is of vital importance therefore, when designing training programs, that the tasks one sets trainees have either a goal of a specific outcome or a goal that is purely pattern search in nature. The mixture of these goals is likely to prevent subjects from learning explicitly. The advantages of explicitly learning rules are numerous. In Experiment 1 it was demonstrated that explicit learning leads to a control performance advantage over instance learning. It was also demonstrated that explicitly learning instances leads to a control performance advantage over the purely implicit learning of instances[4]. In chapter 6 it is demonstrated that explicit rule learning is more transferable than instance learning (implicit or explicit). A caveat to this training programme advice however, is that the results reported here are where subjects are learning non-salient material. The effect of different learning goal instructions may well be different the higher the salience. As mentioned before, the interaction of the levels of salience factor with type of learning goal is something that needs to be examined.

An important and vital element of the task is the prediction questions. It is the prediction questions that allow identification of exactly what sort of learning the subject is using. The key to using the prediction questions as a learning labelling device is by having the different types of prediction question that have been included in the studies presented in this thesis (i.e. Old-wrong, Old-correct, and New types). Most other studies that use dynamic control tasks to explore learning use prediction questions. In all of these studies[5] the prediction questions are created in a very haphazard manner with no definition of whether a prediction question is a New, Old-correct or Old-wrong type. In fact the prediction questions are usually created by arbitrarily choosing the various inputs and outputs that the subjects will have to make their predictions from

[4] This was demonstrated by the specific goal group's performance advantage over the dual goal group during the learning phase. This could alternatively be interpreted as explicit learners of instances learning quicker. However it is interpreted, it still leads to a performance advantage for explicit instance learners.

[5] Studies by Marescaux, Luc and Karnas (1989) and Dienes and Fahey (1995) did use post task questioning that distinguished between New and Old situations. However the post task questions were not the same as prediction questions.

(e.g. Berry and Broadbent, 1984; Berry, 1991; Berry and Broadbent, 1988; Hayes & Broadbent, 1988). Since the vast majority of the prediction questions are likely to be either New or Old-wrong, performance on this sort of prediction questionnaire after doing the learning task with a concurrent secondary task is likely to be at chance level. The prediction questionnaire needs to be created carefully with the important inclusion of a distinct set of Old-correct prediction questions. The data in the study presented here suggest that, with a secondary task, these kinds of prediction questions are the only prediction questions that subjects can perform correctly on. Therefore without the deliberate inclusion of a sizeable amount of this type of prediction question, data will falsely indicate that subjects can only make predictions at chance.

Connected to this, there has been debate as to whether, amongst other measures, prediction questions are a good measure of subjects' knowledge (see Shanks & St. John, 1994, pg. 383). Shanks et al put the question "Can we be certain that the questionnaire procedure exhausts the subject's knowledge of the task?" Specifically, they address concerns that the questionnaire may not be sensitive enough to provide evidence about the true nature of subjects' learning. Shank's point was made about studies that did not use the different types of prediction questions that have been used throughout this thesis. For the studies without the different prediction question types the concern is well founded. The argument in the previous paragraph supports this - If subjects' learning of a dynamic system while performing a secondary task was explored using the arbitrarily created prediction questions, then the results would lead to the erroneous conclusion that subjects can only predict at chance. The method of creating prediction questions presented in this thesis provides a measure of learning that is much more sensitive. The data presented in both this experiment and in Experiment 1 show that these types of prediction questions do not merit Shanks et al's concerns. It is clear from this study that prediction questions can show evidence of subjects' knowledge that results from implicit instance learning. The data from Experiment 1 of the non-specific goal group clearly demonstrates that the prediction questions can show evidence of subjects' knowledge that results from explicit rule learning. The data from the specific goal group reported in Experiment 1, primarily by using the prediction questions, is distinguishable from the explicit rule learning data of the non-specific goal group and the implicit instance learning of the dual goal group. Therefore, these different sorts of prediction questions are sensitive enough to provide evidence for explicit and implicit instance learning and explicit rule learning.

In the introduction of this experiment, it was postulated that for the use of correct instances, the link between instance match and prescribed action, not requiring extra cognitive processing, was direct. It was suggested that this made the use of correct instances very procedural in nature and concluded that their use should be automatic and therefore immune to the effects of working memory load. The results in this study clearly support this. However, there is an alternative and perhaps better way of theorising about why the direct link should lead to the use of correct instances being automatic / implicit. As detailed in the introduction of this experiment, look-up table models match the data of the subjects presented in this study and of the dual goal group in Experiment 1. These models have been designed to explain implicit learning and therefore don't expect any use of working memory in their function. Rather than saying the use of correct instances is like the automatic use of Andersen's (1983) ACT style procedures, it is simpler to explain their automaticity in the terms of the functioning of a look-up table. The problem with using the 'procedures' line of argument is that according to the ACT model, procedures have to start as declarative knowledge and then through the repeated explicit use of this knowledge, they become automatised as procedures. The data from this experiment clearly can not be fitted into that model in its present form. Working memory was so occupied by the secondary task that the repeated explicit coverage of declarative knowledge would not be possible. One could of course enhance the ACT model to allow for some knowledge to be directly stored as procedures. In the light of the large, post ACT, body of implicit learning material, this may well be a necessary and useful task.

The results of this study allow stronger claims to be made as to the implicitness or explicitness of the learning of the different goal groups in Experiment 1. At one end of the scale there is the archetypal explicit rule learning of the non-specific goal group, whereas at the other end of the scale there is the purely implicit instance learning of the dual goal group. In between there is the explicit (and maybe partially implicit - discussed below) instance learning of the specific goal group. The notion of having *three* different forms of learning can help enhance work in other areas of cognitive psychology. Stevenson (1997) used the notion of three different sorts of learning - particularly the idea that there could be a distinction between implicit instance learning using just Old-correct instances and explicit instance learning also using Old-wrong instances - to enhance Evans & Over's model of reasoning (Evans & Over, 1997). Evans & Over had argued for a model of reasoning that consisted of implicit heuristic processes and explicit analytical processes. Stevenson suggests that the results of the data in the present experiment and in Experiment 1 imply that heuristic processes may be purely implicit, however some of them are likely to be explicit as well. Stevenson goes on to suggest that by concluding that some heuristic processes may be explicit as well as implicit it "may help to identify more precisely the kinds of processes that facilitate or impede the ability to [reason logically]".

Finally, the consequences of concluding that a dual goal group leads to purely implicit instance learning must be considered. Firstly, the results of the specific goal group presented in Experiment 1 are examined. It was suggested that this group is learning instances explicitly. The results from the study presented here go some way to supporting this claim. Specific goal subjects without a secondary task can perform well on Old-correct *and* Old-wrong prediction questions. When explicit processes are prevented by a secondary task, subjects' performance is reduced to that of only performing well on Old-correct questions. This implies that they are no longer recording or using Old-wrong instances. Notably, it indicates that some of their learning was explicit (i.e. using working memory) and therefore unlike the dual goal group, some of their instance learning was explicit. What the data cannot resolve is whether control task subjects are learning all instances through explicit processes or just the Old-wrong ones. In other words, was their learning a mixture of implicit and explicit processes or just explicit processes? This is a question that must be answered as with the true mix of implicit and explicit processes of the specific goal group established one can make comparisons with the dual goal group and deduce much information about how implicit and explicit processes interact. The need for the true nature of the learning to be established, so as to allow such useful comparisons is demonstrated in the following paragraph.

As to whether implicit and explicit learning processes occur in parallel, separately, are mutually supportive, or conflicting has been given much debate (see Shanks & St. John,1994; Berry & Dienes, 1993; Seger, 1994). The results of Hayes and Broadbent's study suggested that implicit and explicit processes conflicted with each other, with explicit processes preventing implicit processes. However, failure to replicate their study, suggests caution when regarding this conclusion. The results from Experiment 1 and the data presented here would run contrary to this conclusion. Comparisons with the results presented here and in Experiment 1 demonstrate that the dual goal group's data is the hall mark of implicit instance learning. Notably implicit instance learners can make successful predictions on Old-correct prediction questions. If explicit processes prevented implicit processes then one would expect to find the specific goal group in Experiment 1 (whose explicit processes were still functioning) to show lower performance on the Old-correct instances than the dual goal group who had no explicit processes. The results from Experiment 1 show that both groups are performing comparably on this variable, thus indicating that the explicit processes of the specific goal group are not impeding the learning of Old-correct instances. However, this conclusion is based on the notion that the specific goal group is learning using a mixture of implicit and explicit processes. Specifically, it has not been established whether correct trials are learned implicitly or explicitly for Experiment 1's specific goal group. As mentioned in the previous paragraph this needs to be confirmed to firmly make the conclusion reached in this paragraph.

Another detail that needs to be examined is how the dual goal group subjects process incorrect trials. Are subjects recording these instances and simply not using them when answering the prediction questions or are they completely ignoring them and not memorising them? Subjects' memory from all the goal groups for the various kinds of instances is explored in the next chapter.

In conclusion, the results indicate that the learning process of the dual goal group does not use working memory and therefore can be regarded and labelled as implicit. When the control task, pattern search and dual goal groups were given a concurrent task the results were identical to those of the dual goal group without the concurrent task. It was concluded that the concurrent task of random digit generation sufficiently prevented the use of working memory so that the only processes left were implicitly executed. The use of prediction questions as a labeller of learning processes was reflected on and it was concluded that the Old-wrong, Old-correct and New questions are vital to allow the prediction questionnaire to be a useful and sensitive measure of learning. Finally the consequences of concluding that the dual goal group was learning implicitly were explored. It was suggested that the conclusion opens up a number of lines of reasoning that must be pursued, particularly establishing exactly what mix of implicit and explicit processes the control task group uses. This is important as it would allow questions about how implicit and explicit leaning processes interact to be answered.

Chapter V

Experiment 5: Learning Goals And Memory

The purpose of this study is to test how the different modes of learning induced by the different learning goals affect memory. Considering the suggested effects on learning that these goals should be having, a number of predictions are possible about how subjects should perform on a memory test. Confirmation of these predictions would reinforce the notions about the effects on learning of the different learning goals. The experiment described here is almost identical to Experiment 1. Groups of subjects are given either a pattern search, control task or dual learning goal. Subjects have to learn the Clegg version of the person interaction task consisting of the usual learning phase, test phase and rule description questions. However, replacing the prediction questions is a memory test. The memory test is an instance recognition test where subjects are presented with instances looking similar to the trials they have just performed and asked if they recognise them. The instances consist of either Old (familiar) instances or New (unfamiliar) instances. The Old instances are taken from the trials of the experiment. Some of these trials will be instances when the subject had performed correctly (Old-correct instances) and some when the subject had performed incorrectly (Old-wrong instances). The New instances also fall into two categories. The New instances are generated so that they are different from any that the subject will have seen. One way to generate these instances is so that they follow the rule that underlies the person interaction task (New-legal instances). The other way to generate these instances is so that they flaunt the underlying rule of the person interaction task (New-illegal instances). What follows is a consideration of thinking behind memory recognition tests. Then there is a detailed description and explanation of the predicted results for the three different goal groups. Included in this, relevant literature is described that supports the predictions. Finally, before the experiment itself, the hypotheses are summarised.

There are some important assumptions behind memory recognition tests. Firstly, if subjects have a memory of an instance it is assumed that they can confidently confirm that they have experienced the instance. On the other hand, due to everybody's implicit understanding that memory is imperfect, it is not possible to be quite so confident that an instance has not been seen before. The result of confidence being higher for a memory match than for a non-match is that in tests of recognition, subjects should make more correct recognition judgements when they have a memory of an event than when they do not. The easiest assumption to use is that when subjects do not have a memory of an instance they will simply guess and therefore perform at chance (Baddeley, 1990, pg. 272). Consequently the less memory that a subject has for a specific category of instances, the nearer to chance they should be on this category. Therefore, in this study, all other things considered, subjects performing better on one category of instances than another should be reflecting a stronger memory for that category of instances. This line of thought is most appropriate for performance on Old type instances considered next.

The three learning goals used in this thesis have been said to lead to two broad modes of learning. The pattern search learning goal is said to lead to explicit learning of the underlying rule that underpins the person interaction task. (The predictions about how these subjects should cope on a memory test are dealt with below.) The control task and dual learning goal have been said to lead to instance learning, but of two different forms, that have been described as using two different look-up table models. The control task subjects performed equally well on Old-wrong and Old-correct prediction questions and better on these than on new questions. It was concluded therefore, that these subjects were learning by creating a look-up table into which all instances (both correct and wrong) were entered. The dual goal subjects however, performed significantly better on Old-correct compared to Old-wrong and New prediction questions, and performance on the Old-wrong and new questions was comparable. It was concluded that these subjects were learning by creating a look-up table into which only instances in which subjects performed correctly on were entered. The look-up tables, used to describe instance learning, are essentially supposed to be a representation of what subjects have in their memory. Therefore, the assumption is, that all instances in memory are used to construct the look-up table (Dienes and Fahey, 1995). Considering the two different types of instance learning that are supposed to result from the control task and dual learning goals some clear predictions about subjects' performance on a memory test follow.

Starting with the Old instances; reflecting the dual goal subjects' supposed recording of correct instances only, it would be expected that these subjects should make significantly more 'yes, I have seen that instance before' responses on Old-correct instances than on Old-wrong instances. Reflecting the control task group's supposed recording of all instances, it would be expected that these subjects should make equal amounts of 'yes' responses for Old-wrong as for Old-correct instances.

Now instance learners' performance on the New instances are considered. The New instances, typically referred to as distracters, are included so that subjects do not work their way through the memory test and quickly realise that all the instances they are presented with they have seen before. Instance learners from both the control task and dual goal group should perform similarly on the New instances. For the New-legal instances subjects obviously have no record of the instances to guide recognition. Also, the instances follow the underlying rule of the task and therefore, there should be nothing unusual about them which could hint that they have not been seen before. So, considering the assumptions about recognition tests, both control task and dual goal subjects should perform at chance for the New-legal instances. For the New-illegal instances, again, subjects do not have a record of the instances to guide recognition. However, for these instances, they do not follow the underlying rule of the task and this may act as a hint that the subjects have not seen the instance before. Nearly all studies of artificial grammar learning show that the subjects (who are arguably also performing instance learning), are able to detect instances that are flaunting the rule of the grammar they have just learned (see Reber, 1989, 1993, or Seger, 1994, for an overview of results of artificial grammar learning studies). Therefore it is plausible that instance learners of the person interaction task should be able to detect instances that flaunt the rule that generated the instances they had just experienced. To date, this sort of test has not been carried out on the person interaction task. The results would indicate the extent to which instance learners of the person interaction task can show pattern matching abilities. If subjects are able to detect that they have not seen the New-illegal instances before then this would indicate that instance learners can show pattern matching abilities.

The effect on memory of explicit rule learning of the underlying task is now considered. These are the results that are expected of the pattern search subjects. There is much evidence to suggest that explicit rule learners do not memorise instances, but just store the learnt rules and make recognition judgements on whether or not an instance follows the rule (e.g. Barclay, 1973; Nosofsky, Clark & Shin, 1989). Barclay presented subjects with a string of sentences that described the spatial layout of a group of objects (e.g. "the giraffe is to the right of the lion"). Some subjects were told to memorise the sentences while others were told to figure out the layout of all the objects. In a recognition test memorisers were accurate at identifying the sentences from a mixture of Old sentences and New (distracter) sentences. By contrast, subjects who had to determine the layout of objects identified any sentence that matched the layout as having been seen before. Nosofsky et al studied concept learning. Some subjects while being exposed to the exemplars were given instructions that encouraged rule learning. In a following recognition test these subjects made more 'yes' responses when the instance was consistent with the rules, irrespective of whether or not they had seen the instance before. The rule learners in Nosofsky et al's study and the subjects determining the array of objects in Barclay's study can be compared to the rule learners of this experiment. It is predicted therefore, that they also will only store the rule. Consequently, pattern search learners should have no benefit of a store of instances to make their recognition judgements. Their recognition judgements should be based on whether or not an instance follows the rule they have deduced. This should mean they should recognise as having been seen before all instances that follow the rule. So, subjects should give an equal number of 'yes' responses to Old-wrong, Old-correct *and* New-legal instances, and they should be above chance at rating the New-illegal responses as having not been seen before.

The specific hypotheses: The pattern search subjects: These should shows signs of learning explicit rules. There were no prediction questions in this study, so the main indicator of this should be that they perform well on the rule description questions. They should also be better than the other groups at controlling Clegg during the test phase as was shown in Experiment 1. On the recognition test they should make a similar number of 'yes' responses for the New-legal, Old-wrong and Old-correct instances. The number of 'yes' responses should be above chance for these three instance types. For the New-illegal instances there should be significantly fewer 'yes' responses than that expected by chance, indicating that they are above chance at stating that they have not seen the instance before.

The control task subjects: These should show signs of instance learning. On this experiment this should consist of poor performance on the rule description questions. For the recognition test subjects should give a comparable amount of 'yes' responses on Old-wrong and Old-correct instances (with performance above chance for both these instance types). They should perform at chance for the New-legal instances and below chance for the New-illegal instances indicating that they can tell that they do not conform to the general pattern the trials conformed to.

The dual goal subjects: These should also show the signs of instance learning that are predicted for the control task subjects. For the recognition test subjects should give more 'yes' responses for the Old-correct instances (which should be performed above chance) than for the Old-wrong instances (which should be performed at chance indicating that subjects do not have the instances in memory). For the two New instance types, performance should be the same as that predicted for the control task subjects.

METHOD

Subjects: The 36 volunteer subjects were Durham University graduate and undergraduate students, aged between 18 and 24.

Design: The design was identical to that of Experiment 1 (see pg. 21), except that in place of the prediction questions there was the memory test. As was the case with the prediction questions, the memory test followed the test trials. The unexpected memory test consisted of 60 instances and subjects had to determine whether or not they had seen them before. Following the memory test the rule description questions were presented. There were 4 types of instances in the memory test - 15 new instances that followed the underlying rule of the system, 15 that flaunted the rule, and 30 old instances which were the ones the subjects had engaged in during the test trials. The old instances consisted of some instances that the subject had performed correctly during the test trials - the Old-correct instances, and some which the subject performed incorrectly - the Old-wrong instances. Display of the type of instance was random as was the display of the order of the Old instances.

The Memory Test

Recognition of Instances: There were 60 instances subjects had to categorise as having seen before or not. The recognition of each instance took the following form: First a typical trial situation was presented. The subject's and Clegg's subsequent behaviour was displayed on a clear screen. Subjects then simply had to type 'y' to say they recognised having seen the situation before or 'n' to say that they did not. The instances were generated in four different ways:

A) 'New-legal' instances; Each situation was generated randomly from a list of all *possible* trial situations that the subject had not encountered during either the learning trials or the test trials or earlier during the memory test. The sequence of behaviours was generated so they followed the underlying pattern that Clegg followed.

B) 'New-illegal' instances; The same as the New-legal instances except that the sequence of behaviours was randomly generated from the full set of behaviours that would be *impossible* if they had been created from the underlying pattern that Clegg followed. For instance 'You were *Affectionate*, Clegg was then *Very Rude'* . For this to occur Clegg's previous behaviour would have to have been 8 positions above the *Loving* behaviour. As *Loving* is the highest behaviour this instance could never have occurred.

C) 'Old-wrong' instances; Each instance was randomly selected from all the trials on which the subject had not managed to make Clegg *Friendly*, *Very Friendly* or *Affectionate* on during the test phase.

D) 'Old-correct' situations; Each instance was randomly selected from all the trials on which the subject had managed to make Clegg *Friendly*, *Very Friendly* or *Affectionate* on during the test phase.

15 legal and 15 illegal New type instances were produced. All the 30 Old type instances from the test trials were displayed.

Procedure: Subjects were randomly allocated to one of the three goal groups. As described in Experiment 1 (see pg. 21), the three groups received identical initial instructions apart from one sentence. This sentence dictated the aim of that particular group for the learning trials. Apart from this, the remainder of the experiment was identical for all groups:

The instructions explaining the nature and aim of the subjects' initial learning task were presented first. These were followed by the learning trials. On completion of this phase, subjects from the three groups received identical instructions describing their new goal for the test trials and then the test trials started. Clegg initiated both learning and test trials by displaying one of the three adjectives centred on *Polite*. Following the test trials the subjects were presented with instructions for the memory test. These instructions described the nature of the test and gave an example of an instance from which the subjects would have to make a recognition judgement. After completing the memory test subjects were given a pen and paper and were asked to answer the two rule description questions.

RESULTS

Learning Trials

Trials were scored as correct for the control task and dual goal groups in the same way as that described in Experiment 1. Due to the lack of a specific goal for the pattern search subjects during the learning phase, no measure could be made for their performance during the first set of trials. So only the learning of the control task and dual goal groups could be assessed, since only these two groups had a specific learning goal. The mean number of correct learning trials for these two groups are shown in Figure 5.1. In the Figure, data are shown for all 30 learning trials combined and for each half of the learning trials.

The data in Figure 5.1 were analysed using a 2 (learning goal) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results showed no main effects or significant interactions. This suggests that overall for the learning phase learning goal had no effect on performance. Also, subjects' performance did not change significantly from the first to the second half of the learning phase. (See Appendix 5 for the ANOVA tables and full sets of t-tests for this experiment, pg. 248).

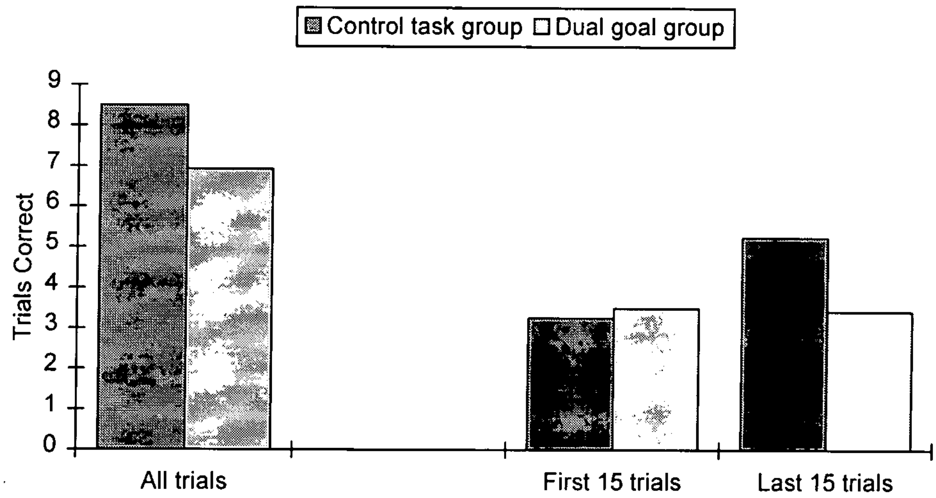


Figure 5.1: Mean number of correct trials in the learning phase for the control task and dual goal groups. Data are shown for all 30 trials combined and for the first and second 15 trials.

The Test Trials

For all three goal groups, correct trials were identified in the same way as for the learning trials. Figure 5.2 shows the mean number of correct test trials for each group for the entire test phase and each half of the test phase.

The results in Figure 5.2 were analysed using a 3 (learning goal) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results showed a main effect of learning goal, $F(2,33) = 6.33$, $p = 0.005$. There were no other significant effects of interactions. To confirm how the groups compared to each other during the test phase between groups comparisons were done on the total test phase scores. The pattern search subjects outperformed the control task subjects, $F(1,22) = 13.54$, $p<0.002$, and just failed to significantly outperform the dual goal group subjects, $F(1,22) = 4.14$, $p<0.06$. The control task subjects and the dual goal group subjects performed comparably.

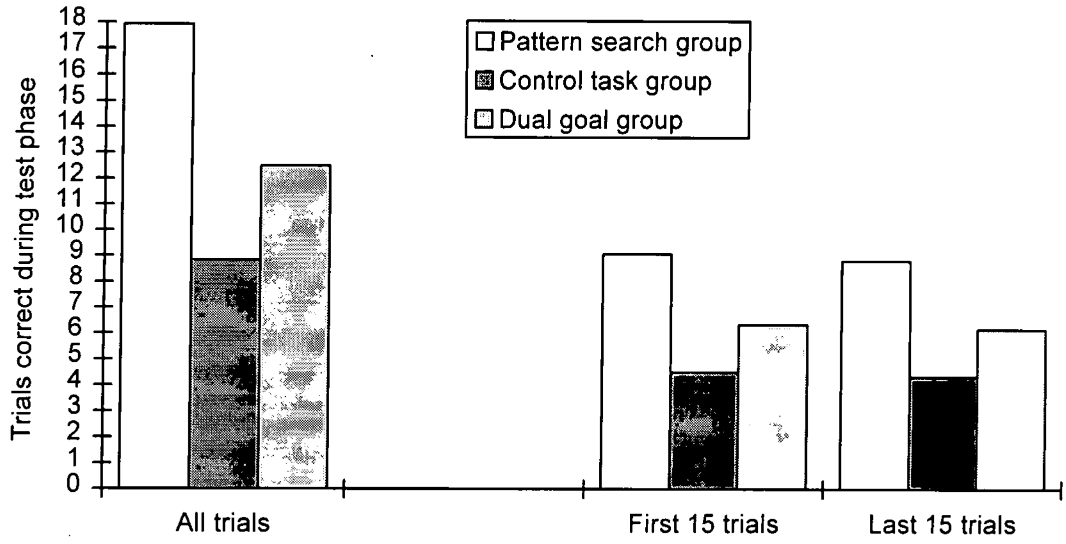


Figure 5.2: Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.

Transfer: In keeping with the rest of the thesis, another important issue to examine is how the specific and dual goal groups coped when the subjects switched goals from the learning phase to the test phase. Two comparisons were made to examine this issue: (i) A comparison between the number of correct trials during the last *half* of the learning phase and the first *half* of the test phase and, (ii) A comparison between the *total* number of correct trials during the learning and test phases. For comparison (i) a 2 (learning goal) by 2 (trial block) analysis of variance with repeated measures on the last factor showed no main effects. The interaction between learning goal and trial block just failed to reach significance, $F(1,22) = 4.19$, $p = 0.053$: the data suggests that there was a tendency for only the dual goal subjects to perform better in the first half of the test phase than in the last half of the learning phase. For comparison (ii) a 2 (learning goal) by 2 (phase score) mixed analysis of variance showed no main effect of learning goal. The main effect of phase score just failed to reach significance, $F(1,22) = 3.68$, $p = 0.068$: there was a tendency for subjects to perform better in the test phase than in the learning phase. The interaction between phase score and learning goal was not significant.

Recognition Of The Instances

For each type of instance, the mean percentage of 'yes' responses for the three groups are shown in Figure 5.3. The total number of Old-wrong or Old correct instances depended on each subjects' performance during the test phase and therefore varied from subject to subject. The results therefore, are displayed as percentages to aid comparisons between the different types of instances. For the Old instances these values represent the percentage of correct recognitions. For the New instances these values represent the percentage of incorrect recognitions.

The data in Figure 5.3 are explored by a 3 (learning goal) by 2 (instance type: New vs Old) analysis of variance with the last factor as a repeated measure. This revealed a main effect of instance type, $F(1,33) = 24.97$, $p < 0.001$. There were no other significant effects or interactions.

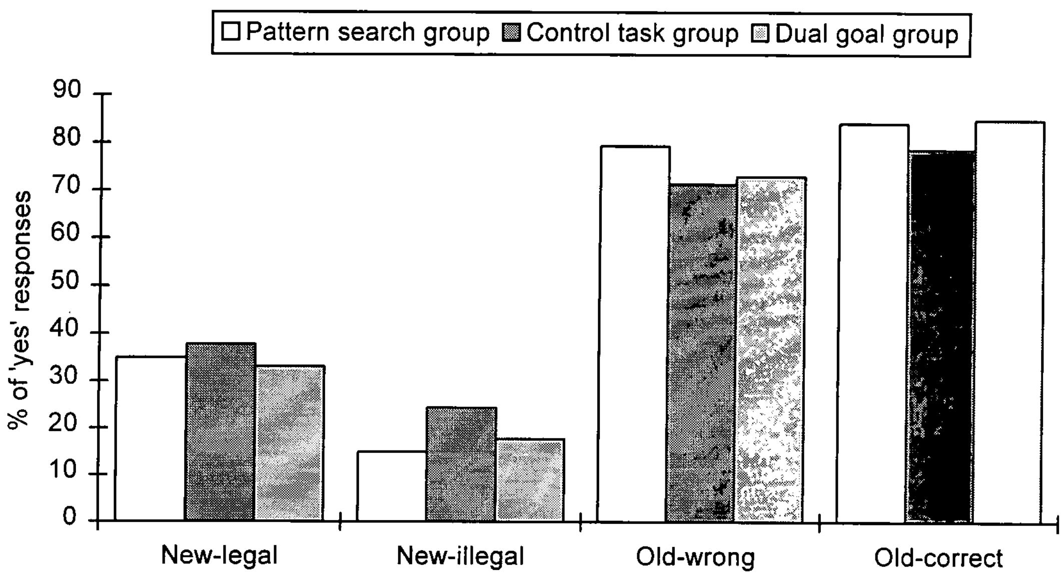


Figure 5.3: Mean percentage of 'yes' answers for the different types of instances.

To examine the effect of instance type, data were collapsed over learning goal and within group comparisons were made. These showed that there was a higher percentage of 'yes' responses to New-legal than to New-illegal instances ($t(22)=5.91$, $p < 0.001$); there was a higher percentage of 'yes' responses to Old-wrong than to New-legal instances ($t(22)=7.96$, $p < 0.001$) or to New-illegal instances, ($t(22)=14$, $p < 0.001$); there was a higher percentage of 'yes' responses to Old-correct than to New-legal instances ($t(22)=9.16$, $p < 0.001$) or to New-illegal instances, ($t(22)=14.13$, $p < 0.001$; and, there was a higher percentage of 'yes' responses to Old-correct than to Old-wrong instances ($t(22)=2.21$, $p = 0.034$).

In the hypotheses to this study it was predicted that there should be some specific within group differences for the instance types. For the pattern search subjects, it was predicted that they should have a similar percentage of 'yes' responses for the Old-correct, Old-wrong and New-legal instance types. Comparison showed that, though there was a similar percentage of 'yes' responses for Old-wrong and Old-correct instances (for, $t(11) < 1$), there was a significantly higher percentage of 'yes' responses for these two instance types compared to New-legal instances (New-legal vs; Old-correct, $t(11) = 6.24$, $p < 0.001$; vs Old-wrong, $t(11) = 5.16$, $p < 0.001$). For the control task subjects, it was predicted that they should have a similar percentage of 'yes' responses for the Old-wrong and Old-correct prediction questions. This was confirmed, $t(11) < 1$. For the dual goal subjects it was predicted that they should have a significantly higher percentage of 'yes' responses for the Old-correct instances compared to the Old-wrong instances. This was confirmed, $t(11) = 2.73$, $p = 0.02$. Notably the results in this paragraph show that the difference in performance between the Old-wrong and Old-correct instances for the data collapsed over learning goal (shown in the last paragraph) is due to this difference occurring for the dual goal subjects as for the other two groups the difference did not occur.

In the hypotheses it was also predicted that for particular groups the percentage of 'yes' responses for some of the instance types should be the same as, around, or on a particular side of that expected by chance (a value of 50%). For the pattern search subjects it was predicted that the percentage of 'yes' responses should be significantly below chance for the New-illegal instances and above chance for the New-legal, Old-wrong and Old-correct instances. Comparisons to chance confirmed these predictions for the New-illegal, $t(11) = 8.68$, $p < 0.001$, Old-wrong, $t(11) = 5.02$, $p < 0.001$ and Old-correct instances, $t(11) = 9.29$, $p < 0.001$, however, the New-legal instances were only marginally lower than chance, $t(11) = 1.91$, $p = 0.082$. For the control task subjects it was predicted that performance should be above chance for Old-wrong and Old-correct instances, at chance for the New-legal instances and below chance for New-illegal instances. Comparisons confirmed these predictions for the Old-wrong, $t(11) = 3.89$, $p = 0.003$, Old-correct, $t(11) = 3.35$, $p = 0.006$, and New-illegal instances, $t(11) = 4.97$, $p < 0.001$, and partially confirmed the prediction for the New-legal instances, the comparison being only marginally lower than chance, $t(11) = 2.19$, $p < 0.06$. For the dual goal subjects it was predicted that performance should be above chance for Old-correct instances, at chance for Old-wrong and New-legal instances and below chance for New-illegal instances. Comparisons confirmed these predictions for the Old-correct, $t(11) = 5.44$, $p < 0.001$ and New-illegal instances, $t(11) = 6.91$, $p < 0.001$, and partially confirmed the prediction for the New-legal instances with them being only marginally lower than chance, $t(11) = 2.16$, $p < 0.06$. However, performance for the Old-wrong instances was significantly above chance, $t(11) = 5.9$, $p < 0.001$. It should be noted that for all groups, the comparison to chance for the New-legal instances showed that they were marginally *below* that expected by chance, indicating that all subjects had a tendency to make correct rejections for these instance types.

The Rule description questions

Subjects' answers to the two rule description questions were ranked in the same way as described in Experiment 1 (see pg. 32). Both judges categorised the answers identically. These categorisations can be seen in Table 5.1.

Table 5.1
Numbers of subjects giving each category of response to the rule description questions

Category / **Group**	No information or Wrong	Partially Correct	Correct
Pattern search	**3**	**1**	**8**
Control task	**10**	**1**	**1**
Dual goal	**8**	**2**	**2**

As can be seen from the data in Table 5.1, the control task and dual goal subjects achieved far less answers in the *Correct* category and far more answers in the *No information or* *wrong* category than the pattern search subjects. Fisher exact probability tests comparing the number of answers in the *No information or wrong* category and in the *Correct* category showed these differences to be highly significant; for pattern search subjects vs control task subjects $p < 0.004$, and for pattern search subjects vs dual goal subjects $p < 0.03$. There was no significant difference between the control task and dual goal subjects. Thus, the pattern search subjects were better than the other two groups at producing answers containing declarative knowledge. As in previous experiments the same comparisons were made again, but with the answers in the *Partially Correct* and *Correct* categories added together, making a stricter criterion for explicit rule learning. Once again the tests showed the same pattern as before.

DISCUSSION

The purpose of this experiment was to examine the effect on memory of the different learning goals, and consequently examine the effects on memory of the different induced modes of learning. As far as can be concluded considering the lack of prediction questions, as in the other experiments of the thesis, the results suggest that the different learning goals have resulted in the different modes of learning. Therefore, the effects on memory of these different modes of learning can be considered to be being examined.

A number of key conclusions can be drawn from the results. (1) As predicted, the control task subjects are performing instance learning by memorising all instances, thus allowing subjects to use both wrong and correct instances in a look-up table. Therefore, their prediction question performance in comparable experiments in this thesis can be considered to be a product of all of their instance memories. (2) The dual goal subjects are performing instance learning by memorising correct instances better than incorrect instances. However, the results suggest that the subjects are also memorising a considerable number of the incorrect (Old-wrong) instances, more than would be expected if a dual goal subject's look-up table is deemed to be a construct of all their instance memories. Therefore, dual goal subjects' prediction question performance in the other comparable experiments in this thesis can not be considered to be a simple reflection of all their instance memories. (3) The idea that pattern search learners are only storing rules and consequently, are only able to make recognition judgements based on whether an instance follows the rule, is not supported by the results. The results suggest that the pattern search subjects learn instances as well as rules. (4) As predicted, both instance learning groups are able to perform pattern matching abilities that allow them to detect when an instance does not fit the pattern that they have just experienced. For the rest of the discussion, firstly, the effect of learning goals on learning during this experiment is examined, then, each of these four key conclusions are considered in detail.

The effect of learning goals on learning

It is concluded that different learning goals result in the different modes of learning from the following points. It is deduced that the pattern search subjects explicitly learned the underlying rule of the task because; (a) as in Experiment 1 the pattern search group performed better during the test phase than the other subjects; (b) the pattern search subjects performed well on the rule description questions with 2/3 of the subjects explicitly describing the rule that Clegg followed. It is deduced that the control task and dual goal subjects learnt through instance learning as the subjects were poor at explicitly stating the underlying rule that Clegg followed with over 3/4 of the subjects not being able to state the rule. Unfortunately, to distinguish between the type of instance learning of the dual goal and control task subjects prediction questions are needed. As the trial situations were already being used in the memory test, it was not possible to have prediction questions. However, the results on the memory test supports the notion that the two groups were learning differently. As predicted and described below, the control task subjects memorised all instances equally whereas the dual goal subjects memorised more Old-correct instances than Old-wrong ones. Another reason to conclude that the three groups learned differently is that the learning goal effect has been seen in other comparable experiments in this thesis - Experiment 1, Experiment 2 for the model subjects (a dual goal was not included), and Experiment 6 (again, not including a dual goal). Therefore it is valid to assume that it occurred here. Consequently, it is possible to fully consider the other results under the assumption that the three goal groups represent the different modes of learning.

The control task group

The conclusion that a control task goal leads to instance learning in which both correct and incorrect instances are memorised equally comes from the following deductions: (a) As predicted, the subjects were equally good at making recognition judgements for Old-wrong instances and Old-correct instances. This implies that their memories for these two kinds of instances are not biased towards correct instances. It also suggests that control task subjects' prediction question performance in the other experiments in this thesis can be considered to be a reflection of all their instances that they have in memory. As their post learning task abilities are described in terms of a look-up table model, it also supports the assumption that their look-up tables are constructed from all the instances they have in memory. (b) Also as predicted, subjects did not perform significantly different from chance for the New-legal instances, further evidence that instance learning, not rule learning, occurred. Performance on the New-illegal instances is considered below.

The dual goal group

The conclusion that the dual goal group is performing instance learning by memorising more Old-correct than Old-wrong instances is deduced from the following point. As predicted, subjects made significantly more 'yes' responses for Old-correct instances than they did for Old-wrong instances. This result also confirms the implicit assumption behind the prediction of 'more 'yes' responses on the Old-correct instances', that dual goal subjects have a stronger focus on correct instances.

The implications of dual goal subjects performing above chance on Old-wrong instances are now discussed. As for the control task group, it was suggested in Experiment 1 that the post learning abilities of the dual goal subjects are a product of their memories that in theoretical terms can be considered to form a look-up table. The automatic assumption is that the look-up table is constructed from *all* their memorised instances. The prediction question results of Experiment 1 and Experiment 4 suggest that the dual goal subjects are able to use Old-correct instances but not Old-wrong instances to make predictions. Due to the assumption that look-up tables consist of all memorised instances, it was assumed that Old-correct instances were the only ones that dual goal subjects memorised. However the results of this study suggest that a significant number of Old-wrong instances are also memorised - subjects perform significantly above chance for the Old-wrong instances. This brings into doubt the assumption that a look up table used to describe a dual goal subject's post learning abilities is constructed from *all* memorised instances. The question is, if the dual goal subjects have a memory of the Old-wrong instances then why are they not also entered into the look-up table? There are a number of explanations that can be put forward to explain the finding that dual goal subjects recognise Old-wrong instances at above chance level.

One explanation exists that does not try to explain why Old-wrong instances are not entered into the look-up table. For this explanation the very idea that Old-wrong instances are not entered into a look-up table is challenged. It could be that dual goal subjects are entering both Old-wrong and Old-correct instances into the look-up table and their performance on prediction questions is just a reflection of the relative strengths of their memory on the two instance types. For this argument to stand up, one would expect to find the *raw* strength of the memories for the Old-correct instances to be stronger than that for the Old-wrong instance type. In other words you would expect to find literally more correct instances in the look-up table, not just a higher percentage. This prediction can be tested in the results of this experiment. The memory test used in this experiment was a recognition test. Therefore, to calculate raw strength of memory you can not simply see how many Old-wrong and Old-correct instances the subjects got right, but you must also take into account when subjects were simply guessing. So, to explore this prediction, a raw measure of the strengths of subjects' memories for the two instance categories can be considered to be the number of instances subjects correctly recognised (their hits) minus the number of instances they incorrectly recognised (their false alarms). As predicted by this explanation the mean hits minus false alarms for the Old-correct instances was larger than that compared to the Old-wrong instances [10 (11.25 - 1.25) compared to 7.16 (12.33 - 5.17)].

There is a caveat to this argument. It suggests that performance for the Old-wrong prediction questions should be above chance as subjects should have some Old-wrong instances in the look-up table, but just not as many as Old-correct instances. The evidence for this presumption is mixed. Contrary to this presumption, for Experiment 1, performance for the dual goal subjects for the Old-wrong prediction questions was not above chance (a value of 20%), $Z = -1.59$, $p = 0.11$. However, although not quite comparable, the data from Experiment 4 has a larger sample size (72 vs 24) and performance on the Old-wrong prediction questions was above chance (a value of 20%), $Z = -6.25$, $p < 0.001$.

The second explanation concerning the good memory for Old-wrong instances does focus on why dual goal subjects do not enter Old-wrong instances into a look-up table. This is that subjects' memories for each individual Old-wrong instance may have not been as strong as subjects' memories for each Old-correct instance. Perhaps instances are only entered into a look-up table when subjects have a strong memory for them. It is plausible to argue that Old-correct instances are going to be focused upon more strongly and thus lead to a stronger memory than Old-wrong instances. The results already indicate that memory for them was better. Also, in Experiment 4 (see pg. 105), it was explained that Old-correct instances have a more direct link with action than Old-wrong instances. It is possible that this direct link with action, which does not occur for the Old-wrong instances, could lead to the Old-correct instances being encoded more strongly. An alternative way to look at this strength of memory explanation is rather than saying it is the strength of the memory of the instance that allows it to enter the look-up table, it could be the nature of the memory. Perhaps the direct action link of Old-correct instances is the specific quality that leads to it being entered into the look-up table. Why this direct action link is not required for the control task subjects could be because the subjects are not learning wholly implicity and therefore a direct action link is not so important. So, in summary this explanation concedes that Old-wrong instances are not entered into a look-up table, and suggests that the memory for these instances may be either quantitatively different (not as strong) or qualitatively different (not closely linked to actions) and therefore for these reasons the Old-wrong instances are not entered into a look-up table.

Finally, another explanation that may explain dual goal subjects good performance on Old-wrong instances is one based on a potential methodological flaw. An Old-wrong instance was any instance in which Clegg made a response other than *Friendly, Very Friendly* or *Affectionate*. This includes instances where Clegg ended up being *Indifferent, Polite* or *Very Polite*. These were the responses that meant that the instance would have been correct if it had been taken from the learning phase. It is possible that an Old-wrong instance taken from the test phase could be identical to a correct instance from the learning phase. Therefore, dual goal group subjects may be scoring above chance for Old-wrong instances due to an identical instance being memorised during the learning phase because then it fell into the Old-correct category. If this argument was shown to be true, then it would indicate that Old-wrong instances were not really being recognised above chance as they had been memorised, but because Old-correct instances of the learning phase were being used to help recognise them. The results for all the groups were reanalysed with these Old-wrong instances that were identical to the Old-correct instances from the learning phase removed (8% of the data). (The results can be seen in Appendix 5, pg. 249.) The results showed an identical pattern to that shown before. The percentage of corrected Old-wrong instances that had 'yes' responses was almost identical to the uncorrected Old-wrong instances (for all the groups there was only a difference of 0.13% and specifically for the dual goal group there was a difference of only 0.29%). These additional analyses refutes the argument that the Old-wrong instances may be above chance due to a problem with the way the instances are categorised.

The pattern search group

The idea that pattern search learners are only storing rules and consequently, are only able to make recognition judgements based on whether an instance follows the rule, is not supported by the results. If this was the case then the pattern search subjects should have given a similar number of 'yes' responses to the New-legal instances as they did to the Old-wrong and Old-correct instances. This did not happen. Subjects got significantly more 'yes' responses for the two Old instance types. Also, again differing from the Old instance types, the number of 'yes' responses for the New-legal instances was not significantly different from that expected by chance. If anything, it was marginally below that expected by chance. In fact the results of the instance recognition test for the pattern search subjects matched that of the control task subjects.

At first sight then, the results of this experiment would suggest that the pattern search learners were learning instances. However, the results of the rule description questions showed that the pattern search learners are definitely learning rules - 2/3 of them could state the rule. The results of the prediction questions in comparable experiments in this thesis (1,3 & 6) also show that subjects are learning rules - they can make accurate predictions from novel situations. One step towards explaining the results is the possibility that pattern search subjects are learning both instances and rules. As suggested by Nosofsky et al ".... mixed models are possible which assume rule induction together with some form of residual exemplar storage (Nosofsky et al, 1989, pg. 285)". If this were the case, then it would explain how pattern search subjects performed well on the Old instances. One possible reason why pattern search subjects may have memories of the instances is that the instances tested came from the test phase where all subjects had a specific control goal. That is, pattern search subjects may have learned rules in the learning phase where there was no specific goal, and then learned instances in the test phase where there was a specific goal. The memory for instances may not have occurred if the Old instances on the recognition test came from the learning phase where there was only the pattern search goal. With the recognition test designed in this sort of way, the situation would be much more similar to the other studies that have shown a lack of memory for instances by rule learners. These studies generally apply their recognition test straight after the learning phase without giving subjects any other task in which they may have been encouraged to memorise instances.

The question still remains though, as to why the subjects did not apply the rule to the New-legal instances and thus give above chance amounts of 'yes' responses. The good performance on the Old instances suggests that subjects may have some memory of instances. One possibility then, is that subjects simply relied on their memory of instances to perform the recognition test. Therefore, pattern search subjects were left to guess about the New-legal instances.

Another possible reason why the rule was not applied to the New-legal instances thus falsely identifying them as having been seen before, is that subjects were not given enough information in the instances to easily apply their explicit knowledge of the rule. When the instances were presented, subjects were only given the subjects' input and Clegg's resulting output. To decide whether the instance followed the rule, subjects had to calculate what Clegg's *previous* output should have been. If that output fell within the limits of the behaviour scale then the instance followed the rule, if it fell outside the limits then it violated the rule. If instances had consisted of this extra item of information then the fact that the instances followed the underlying rule would have been more distinct and so might have allowed the subjects to falsely recognise the New-legal instances. Indeed, one subject (not included in the results) who already knew the rule performed the experiment and tried to classify the instances in the recognition test by applying the rule. Surprisingly, this subject produced a pattern of results identical to the present pattern search subjects. This suggests that the subjects did have difficulty applying the rule to the instances and supports the idea that with extra information in the instance the subjects may be more likely to apply the rule and so wrongly identify New-legal instances as having been seen before.

The New-illegal instances

All three groups were able to detect correctly that they had not seen the New-illegal instances before. Also, all subjects were significantly better at correctly rejecting New-illegal instances than they were at correctly rejecting New-legal instances. These two points are deduced from the facts that for all subjects (a) The number of 'yes' responses were significantly *below* that expected by chance for the New-illegal instances, (b) there were significantly more 'yes' responses for the New-legal than New-illegal instances. Therefore, the fact that the New-illegal instances did not follow the same pattern as the trials must have been detectable by all subjects. It is probable that all three learning groups were using a similar mechanism. It had been predicted that the pattern search subjects should be able to correctly recognise the New-illegal instances by detecting that the instance did not follow the rule they had learned. However as discussed above, it appears that pattern search subjects were not using their knowledge of the rule to make recognition judgements. Therefore it is safe to assume that they were using the same principle as the instance learning groups. The fact that the instance learners were able to detect the illegality of the New-illegal instances is similar to the findings of artificial grammar learning studies (e.g. Whittlesea and Dorken, 1993). It suggests that the mechanisms of instance learners of dynamic control tasks share similarities with instance learners of the more abstract artificial grammar learning.

Exactly what mechanisms are involved in the pattern matching abilities of the subjects in this experiment are debatable. Whittlesea and Dorken, showed that given the appropriate instructions (essentially those that lead subjects to be attentive to the structure of the instances) subjects were able to learn abstract knowledge about an artificial grammar. Proof of this came from subjects' ability at detecting, above chance, novel instances that matched the pattern (or grammar) from the one they had experienced during their training session. The notion proposed was that subjects had some abstract representation of their learning instances that allowed them to perform pattern matching abilities. The key point about an 'abstract representation' is that it implies that more than just instances are stored. Alternatively, in the use of the instances, it implies that more is happening than just the recall of the instances from memory. For the subjects in this experiment, exactly what else is stored (if anything) or what other mechanisms are being used to allow the store of instances to be used to detect the illegality of the New-illegal instances is open to debate.

One possible pointer to solving this debate is by looking at a feature of both New-legal and New-illegal instances. The main distinguishing feature about the illegal instances was that the difference along the behaviour scale between input and output was generally very large. (Since the recognition instance was only made up of an input and an output, this was the only way to generate instances to make them illegal.) It is possible that this was what was detected to identify the illegality of the New-illegal instances. If a simple guiding principle such as this was used then it does not require any more information to be stored in the look-up table than the instances themselves. From a theoretical view point, this means that mechanisms governing the entering and reviewing of instances into the look-up table do not need to be enhanced to take into account the pattern matching abilities of the instance learners.

In favour of subjects simply using a guiding principle to categorise their recognition, features and performances on the New-legal instances should be considered. The creation of the New-legal instances would have led to them sharing similar features with the New-illegal instances such as a large gap between input and output. Another feature that they would have shared is a tendency for the input or output to be towards the *Very Rude* end of the scale. This is because to achieve the test phase goal, the inputs and outputs would have a tendency not to come from that end of the scale. Hence to make an instance novel there would be a tendency for the inputs and/or outputs to have consisted of these tell tale behaviours. If subjects were using some mechanism such as a simple guiding principle that focused on one of these features then this would also explain why subjects had a tendency (the results were only marginally significant) to correctly reject New-legal instances as well as New-illegal instances. The tell tale features would not have been as extreme or common for the New-legal instances which would explain why the result was only marginally significant for the New-legal instances and why more New-illegal instances were correctly rejected than New-legal ones. The important point about this 'guiding principle' argument is that it does not require the instances to have any other information stored than their raw perceptual features.

In summary, the results suggested that the different groups learned differently therefore, as intended, the experiment was able to examine the effects on memory of the different modes of learning. It was concluded that the control task subjects were memorising, both Old-wrong and Old-correct instances equally. This supported the notion that prediction question performance in the other experiments is a reflection of subjects' memories, in which, in theoretical terms, all memories are entered into a look-up table.

The dual goal subjects as predicted, were memorising more Old-correct instances than Old-wrong instances. However, contrary to predictions, they were memorising some Old-wrong instances. This refuted the notion that their look-up tables are composed of all the instances that are in their memories. It also posed the question of how to explain dual goal subjects' prediction question performance in the other experiments of this thesis. A number of explanations were put forward to explain these issues. The first explanation concentrated on explaining the latter problem. The explanation was that prediction question performance could be considered a reflection of all instances in memory and that the better performance of the Old-correct prediction questions was simply due to more Old-correct instances having been memorised. The evidence for this argument was inconclusive. The second explanation turned to the look-up table construction problem and concentrated on explaining how a look-up table model may still be used to explain subjects' post learning task abilities. To do this the assumption must be dropped that a look-up table consists of all instance memories. The explanation suggested that for the dual goal subjects a qualitative or quantitative difference between the Old-correct and Old-wrong instances meant that only Old-correct instances were entered into the look-up table. The Old-correct instances' direct link with action, as described in Experiment 4, was proposed as a feature that causes the qualitative or quantitative difference between Old instance types for the dual goal subjects. One explanation of the memory for Old-wrong instances that was ruled out, was that, the above chance performance on the Old-wrong instances was a methodological artefact of Old-wrong instances actually consisting of learning phase Old-correct instances.

The rule descriptions of the pattern search group confirmed that they were learning rules however the recognition data suggested that they had also memorised instances. It appeared that their memory of instances was used to make their recognition judgements, not their knowledge of the rule of the task. One reason put forward for this was that there was not enough information in the instance to easily apply the rule.

Finally, subjects' performance on the New-illegal instances was discussed. All subjects were able to correctly detect the New-illegal instances as not having been seen before. It was suggested that this may be due to them using a simple guiding principle. If this were true, it would not necessarily require subjects to encode their memories of the individual instances with any extra information than the perceptual features of the instance. This would mean that that the encoding process of the look-up table model would not need to be enhanced.

Chapter VI

Experiments 6a & 6b: Rules / Instances Learning Distinction And Transfer Of Learning

The studies reported in the previous chapters have focused on the different learning goals - control task and pattern search - and the relation of these goals to subsequent learning. Experiment 1 demonstrated that it is the control task goal of a typical implicit learning task that leads to instance learning whereas if given a pattern search goal and asked to learn the same material, the result is explicit rule learning. Amongst other things, the importance of this finding is that two groups of subjects can be given an identical task and one group of subjects can be led to learn the task through instance learning and the other through explicit hypothesis testing and rule learning. Previously researchers had to set different tasks to subject groups in order to observe these different forms of learning. Differences in instance and rule learning therefore, could not easily be explored as differences in performance were arguably due to the differences in the tasks. By simply making the learning goal either control oriented or pattern search oriented on the same task, the qualitative differences of instance and rule learning can be explored.

In the introduction to experiment 1 it was noted that, recently, Shanks and St. John argued against the implicit-explicit learning distinction and proposed that implicit learning is really instance learning. If this is the case, then one would expect that the learning gained in an implicit learning task would not transfer to a semantically dissimilar task. The aim of this study is to test this proposition using the learning goal paradigm. Either instance learning or rule learning is induced by varying the learning goal in the Clegg version of an implicit learning task. The results show that rule learning leads to transfer to a semantically dissimilar task while instance learning shows little or no transfer. In the following sections, the research on transfer is discussed. Then the work on instance learning is returned to and discussed in the light of the work on transfer.

The results from the earlier studies of this thesis could be regarded as showing that instance learning and rule learning differ in the degree to which they support transfer to a novel situation. Instance learners (subjects given a specific goal) could transfer what they had learned to the same task with a different specific goal and to prediction tests where the prediction situation was familiar. (Subjects' performance did not drop significantly overall between learning and test phases, and subjects performed comparably on Old prediction question types and above that of New type questions.) However, they could not transfer what they had learned to prediction situations that were novel or to answering an explicit question about the rule underlying Clegg's behaviour. By contrast, rule learners (subjects given a non-specific, pattern search goal) could transfer what they had learned to all the above situations. To assess this notion of differential transfer as a function of goal specificity, the research on transfer is reviewed.

Transfer of Learning

The question of transfer has a long history in Psychology and it is a topic that provokes extreme opinions, some claiming that it is a rare (Thorndike, 1913) or non-existent (Detterman, 1993) commodity, others that it is ubiquitous in human learning (Ferguson, 1956; Hebb, 1949). In a classic paper, Thorndike and Woodworth (1901) argued that one can expect transfer between two stimulus environments to the extent that they share "identical elements" in common. Despite considerable ambiguity over what constituted identical elements, it was taken to mean identical at the level of surface features. According to this view, if two situations share an underlying deep structure but differ in their surface appearance, transfer cannot be expected, whereas if there are surface elements in common, transfer will occur. Modern proponents of this view include Singley and Anderson (1989), for whom the elements over which surface similarity is defined are production rules, Logan (1988), for whom the similar elements are specific learning instances, and Bassok, Wu and Olseth (1995), for whom similar elements are propositions. Transfer based on surface similarity is usually referred to as near transfer.

The counterclaim in support of transfer was made by Judd (1908) who argued that when learning can be organised around a guiding structural principle, transfer is determined by the degree to which the subject has grasped that principle, through either discovery or instruction. A classic study of principled transfer was conducted by Scholchow and Judd (1898, reported in Judd, 1908). Twelve year old boys were asked to throw darts at an underwater target, a skill that requires considering the deflection that the light suffers through refraction. Half the subjects were instructed in the principle of refraction, the remainder were not. Both groups did equally well at first, since all needed time to practise the skill. But when the amount of water was reduced, thereby altering the degree of deflection of the light, the boys without the principle became confused; practise in one setting did not transfer to the other. By contrast, the boys with the principle adapted readily. The same pattern recurred when the pattern changed again, with the principle group adjusting rapidly over time, and the non principle group adjusting less rapidly. Thus transfer is not automatic but depends upon insight into general principles. Modern proponents of this view include Bassok and Holyoak (1993), Brown (1990) and Novick (1990). Transfer based on deep structural principles despite surface dissimilarity is usually referred to as far transfer.

Near Transfer Based On Surface Similarity

The procedural version of the surface similarity view is based on Anderson's (1987) theory of skill acquisition. According to this theory, the *acquisition of cognitive skills comes about* when declarative knowledge about a domain is converted into procedural knowledge for smooth, fast and accurate problem solving. In the theory, the process of conversion is achieved by using general *purpose weak methods* that can convert declarative knowledge into domain-specific procedures. For example, when programming in LISP, one production rule might be "If the goal is to form a new list from two elements then generate the function LIST with the first argument corresponding to the first element and the second argument corresponding to the second element." These procedures are triggered under highly specific conditions, so that a procedure learnt in one part of a domain will not be used in another part of the domain unless the conditions of use of the procedure, that is the goal in the "if" part of the rule, are identical in the two cases. The theory predicts that there will be little or no transfer between subskills within a complex skill domain when knowledge is used in different ways, even though the subskills might rest on a common body of knowledge.

A number of studies support the idea that transfer can be predicted as a function of overlapping production rules. In one such study, Singley and Anderson (1985) investigated transfer of text editing skill across text editors (see also Bovair, Kieras and Polson, 1990; Polson and Kieras, 1985). Prior task analyses suggested that two line editors showed extensive production overlap with each other but low production overlap with a third screen editor. Subjects were trained to use one of the three editors and transfer to another editor was measured. As predicted, substantial positive transfer was found between the two line editors, which shared a large number of production rules. Also consistent with the predictions, less transfer was obtained between the screen editor and the two line editors. Comparable evidence exists for learning to operate a device (Kieras and Bovair, 1986) and for subjects learning computer languages (Katz, 1991; Wu and Anderson, 1991; Scholtz and Wiedenbeck, 1990).

The procedural theory also predicts that there should be no transfer when the conditions for using the production rules are different, that is, when the goal structures are different, even though they are based on the same declarative knowledge. Three studies appear to confirm this prediction. No transfer was found between the generation and justification of geometry proofs (Neves and Anderson, 1981) or between evaluating to generating LISP programming instructions (McKendree and Anderson, 1987; Kessler, 1988). However, Pennington, Nicolich and Rahm (1995) dispute this claim. In a well-controlled study, they found considerable transfer between generating and evaluating LISP programming instructions. Inspection of subjects' verbal protocols suggested that this was because there was considerable elaboration of declarative knowledge during the learning phase. Thus, there is evidence that transfer does involve more than surface similarity.

The instance view of transfer (Logan, 1988) makes the fundamental assumption that people performing a task store instances of past performance in memory and that each instance is stored as an independent copy or "exemplar". On their first encounter with the task, having no stored instance, people will use whatever strategic, rule-based tools they have available; this constitutes a task "algorithm". Subsequently, however, they will have available not only the algorithm, but also memory of the past instances of performance. When the task recurs, performance is based on the first solution that is retrieved from memory, the algorithm or a retrieved instance. The time to retrieve each past instance is assumed to vary stochastically (the probability being a function of the previous instances) so that the algorithm competes for retrieval with that instance having the fastest current retrieval time, that is, the lowest value drawn from a set with similar distributions, one for each instance that has been stored. With enough stored instances, the algorithm will tend not to be retrieved faster than the fastest instance and so responses will come to rely virtually exclusively on past instances. Automaticity, according to the instance model, corresponds to this shift from algorithmic to instance-based retrieval.

A principal assumption of the instance model is that if a previously solved task is presented, past instances of that same task are retrieved from memory. Thus, learning is item-specific, and retrieval is of the same items that were previously used in the task. It is assumed that transfer between distinct items within the same task does not occur because presentation of a novel item does not lead to retrieval of an item previously used in training. In support of this view, Logan and Klapp (1992) had subjects solve alphabetic arithmetic problems (e.g. if A=1, B=2, etc., does A+2=D?) with one set of ten letters and the digits 2-5 for 12 sessions of nearly 500 trials each. Initially, the subjects' response times increased markedly with a new digit (e.g. there was a longer time for B+5 than for B+2), as if they were moving forward through the alphabet from the given letter for the required number of digits (e.g. for B=2: B, C, D). By the 12th session, the slope of this increasing function had considerably decreased, suggesting the development of a new strategy. The subjects were then transferred on session 13 to a new set of 10 letters. Although the task remained unchanged, the slope of the function relating response time to new digit increased dramatically - to nearly the value it had had in the initial sessions. That is, transfer was extremely limited.

Taken together, these studies provide strong evidence for near transfer based on surface similarity, whether the similarity be defined over production rules or memorized instances. The results of the control task group in Experiment 1 also demonstrate near transfer based on surface similarity. These subjects learned to control the computer person by making him consistently produce a specific response. When the task was changed to a prediction task, in which subjects had to predict Clegg's next response from a sequence of three prior responses, the specific goal subjects made accurate predictions when the prediction sequence was familiar but not when the sequence was novel. Such a result is consistent with the idea that the subjects are retrieving specific memories of learning instances. Control task goal subjects could also transfer what they had learned to a novel specific goal, for example, from making Clegg *Polite* to making him *Very Friendly.* This result would not be predicted by either Anderson's procedural model or Logan's instance model, since the novel goal means that the memorized instances were not the same as those needed for the new goal. However, it is likely that during transfer, subjects initially used their task algorithms to meet the new goal, consistent with Logan's model, at least until their store of specific instances could be used as well. Both the procedural model and the instance model seem able to explain the above results on implicit learning. However, the instance model is generally preferred (e.g. Dienes and Fahey, 1995) and it also gives the best account of performance with a novel specific goal.

Far Transfer Based on Structural Principles

Near transfer based on surface similarity is very well-documented. However, there is increasing evidence for far transfer based on structural principles. Far transfer is involved whenever the learning task and the transfer task differ in surface similarity but have the same underlying structure. Such transfer depends on whether or not the learner has grasped the underlying principle in the learning phase. It is observed in situations in which either the learner is given the underlying principle, as in Scholchow and Judd's (1898) study of throwing darts underwater, or the learner abstracts the underlying principle from the examples during learning (e.g. Gick and Holyoak, 1983; Owen and Sweller, 1985; Sweller, Mawer and Ward, 1983; Vollmeyer, Burns and Holyoak, 1996).

Of relevance to the present study is the work of Sweller and his colleagues (e.g. Mawer and Sweller, 1982; Owen and Sweller, 1985; Sweller, 1988). This work was partly reviewed in chapter 1, however is covered here from a different perspective. In several studies, these researchers have shown that people with a pattern search goal gained more structural knowledge about a task than did people with a specific goal. For example, Owen & Sweller (1985) taught school children some of the basic principles of trigonometry (cosine, sine, tangent). They then asked separate groups of these children to practise on sets of problems with goals differing in specificity. One group (the control task group) had to solve problems with typical, specific goals (e.g. find the length AB on triangle ABC and also find angle A). The other group (the pattern search group) were given a similar triangle and simply told to find as many unknowns as possible. The problems were arranged so that both groups were calculating the same number of sides and angles. The pattern search group clearly learnt better than the control task group: They made fewer mistakes during a testing phase and they showed superior transfer to novel problems.

These studies, using explicit learning tasks, highlight the importance of a pattern search goal for inducing the underlying structural principles of a task and so enabling far transfer to occur. Experiment 1, used an implicit learning task, and showed comparable effects. Subjects who had a pattern search goal showed superior performance across the board in comparison to the specific goal subjects. They showed greater transfer to a novel specific goal, they made accurate predictions from both familiar and novel situations, and, unlike the specific goal subjects, they could state the rule explicitly when asked. However, Experiment 1 was not specifically designed to study transfer. The aim of the study described here is to provide such an investigation. Under investigation is the idea that whether or not far transfer will be seen depends on whether the subject has a non-specific, pattern search learning goal or a specific, control task learning goal.

The Present Study

Previous studies have attempted to show far transfer in an implicit learning task. The attempt appears to have been successful in grammar learning but not in a computer control task. Subjects who are trained on an artificial grammar using one set of letters perform well on test items generated by structurally isomorphic grammars but with different letters (Mathews, Buss, Stanley, Blanchard-Fields, Cho and Druhan, 1989; Reber, 1969). However, it is not clear what mechanism underlies such transfer. Classification performance in artificial-grammar tasks can be accounted for by a variety of mechanisms that do not involve induction of the underlying grammatical rule. These mechanisms include: subjects' ability to indicate grammatical or ungrammatical parts of letter strings (Dulaney et al., 1984); knowledge of bigrams or trigrams

(Perruchet and Pacteau, 1990, 1991; Perruchet, Gallego and Pacteau, 1992); knowledge of chunks (Servan-Schrieber and Anderson, 1990); or knowledge of sequential letter dependencies, that is, the ability to decide, when presented with a string of letters, whether each letter in the grammar, if presented next, would create a grammatical string (Dienes, Broadbent and Berry, 1991). While such fragmentary knowledge is arguably abstract to some degree (Matthews, 1990) it still falls far short of knowledge of the underlying grammatical rule. Indeed, Whittlesea and Dorken (1993) have shown that such abstract knowledge can be learned purely through instance learning. Hence transfer in an artificial-grammar learning task could be due to similarity of instances.

Berry and Broadbent (1988) trained subjects on one dynamic control task and then measured their performance on a second task involving the same underlying rule. The semantic cover story of the transfer task was either superficially similar or dissimilar to that of the learning task. Subjects in the superficially similar condition improved their performance as much as control subjects (who continued to perform the same task). However, subjects in the superficially dissimilar condition showed no such transfer. Subjects who were given a hint that the underlying equation in the transfer task was the same as that in the learning task were not helped. In fact, subjects in the similar condition performed worse when given such a hint. By contrast, such a hint generally aids transfer in explicit tasks involving analogical transfer across semantic domains. In the latter case, subjects seem only to need reminding of knowledge available to them (e.g. Gick and Holyoak, 1980).

Squire and Frambach (1990) trained subjects in a manner almost identical to the condition of the control task subjects presented in the study shown here. Subjects with a specific goal were first trained on Berry & Broadbent's (1984) sugar production task then they had to attempt the Clegg version of the person interaction task. Transfer in the other direction was not examined. It was found that the control subjects did not transfer their learning to the person interaction task. There are a number of differences in the design used in the study reported here that makes the inclusion of the control task subjects necessary and indeed means the study as a whole still serves an original purpose. For one thing, the focus of the Squire and Frambach's study was on amnesiac patients. Also the aim of the study reported in this chapter is to examine the difference of transfer ability between control task subjects and pattern search subjects, something that has not been done to date.

Thus, while there is no clear evidence for far transfer in implicit learning tasks, there is direct evidence for the lack of such transfer in a dynamic control task. In the present study, a similar method is used to Berry and Broadbent's (1988), although no hint is provided to any of the subjects. In agreement with Berry and Broadbent and Squire and Frambach (1990), it is predicted that subjects given a specific, control task goal will not be able to transfer their knowledge to a dissimilar task whether transfer is to the person interaction task as in the Squire and Frambach paper, or from the factory task to the person interaction task, as tested in one of the conditions presented here. However, it is also predicted that subjects given a pattern search goal will transfer their knowledge regardless of whether the learning and transfer tasks are similar or dissimilar.

The critical results expected in this experiment: Initially, there should be evidence that the group with a non-specific, pattern search goal learns through explicit hypothesis testing and rule deduction and that the group with a specific control task goal learns through implicit instance learning. Therefore, for the pattern search group there should be a positive correlation between control performance and prediction question performance. Prediction questions that have been generated from both familiar trial situations and novel situations should be answered comparably. Subjects should be able to describe explicitly the underlying rule that the system is based upon. Conversely, for the control task group there should be no positive correlation between control performance and prediction question performance, subjects should be particularly poor at making predictions from novel situations, and subjects should be very poor at explicitly describing the rule that the system is based on.

With the above criteria met, the main hypothesis of this experiment can be explored. That is, that explicit rule learners can demonstrate transfer whereas instance learners can not. Therefore, the group with a non-specific pattern search goal should be expected to exhibit transfer by showing no overall drop in performance between the first and second tasks, and a significant positive correlation between the first and second tasks. The group with a specific control task learning goal should be expected to exhibit a lack of transfer by showing a significant drop in overall performance between the first and second tasks and no significant positive correlation between the first and second tasks. To strengthen the results, during the second task there should also be significantly better performance from the transferring group when compared to the non transferring group. Finally, the transferring group should exhibit a greater recognition of the underlying similarity of the two tasks than the non transferring group.

The experiment reported below (Experiment 6a) is followed by a second experiment (Experiment 6b). The second experiment was included to examine transfer when the task that subjects transfer to is 100% structurally similar to their initial task. The rationale for this second experiment will be described in the discussion of Experiment 6a.

Experiment 6a - METHOD

Subjects: The 48 volunteer subjects were Durham University graduate and undergraduate students, aged between 18 and 24.

Design: A 2 (goal: specific, control task vs non-specific, pattern search) by 2 (task: person interaction by factory) independent groups design was used. There were two main groups of subjects - a control task group and a pattern search group. Within each group, half the subjects had one task first (the person interaction task) while the other group had the other task first (the factory task).

Subjects completed 30 learning followed by 30 test trials. They were then given the unexpected questionnaire, consisting of 15 prediction questions followed by two rule description questions. There were 3 different types of prediction question - 5 based on old correct trials, 5 on old incorrect trials and 5 on new trials. These 15 questions were displayed in a random order. The order of the two rule description questions was counterbalanced across subjects. Following this subjects had to complete 30 trials on the new task. This was the transfer phase. Finally they received 3 questions relating to how similar they perceived various aspects of the two tasks to be.

Task A; The person interaction task : This was the Clegg version of the person interaction task identical to that used in Experiment 1 (see pg. 21).

Task B; The factory task : Subjects were asked to imagine that they were in charge of a sugar production factory in an underdeveloped country. They could control the rate of sugar production simply by changing the size of the work force, ignoring all other factors. The size of the work force could be one of 12 things; *100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, or 1200* employees. To tell the computer what size they wanted to set the work force to, they simply had to type in a number from 1 to 12 representing the number of hundreds of workers they wished to employ. The factory would start off producing a certain tonnage of sugar. Its output was reported in thousands of tons and could be anything from *1000* to *12000* tons. Once subjects had been told how much output the sugar factory was making, they would then enter in the next size they wanted the work force to be. Once subjects had set the new work force size the resulting factory output would be displayed. Then subjects again altered the size of the work force, and so on. The output size was calculated by the same equation that predicted Clegg's next response (for the equation see Experiment 1 pg. 22). The entire task was carried out using an IBM compatible computer. The possible work force sizes were displayed on a piece of paper attached to the bottom of the screen for permanent reference.

For the learning phase the groups of subjects were given identical instructions except for one sentence. Depending on their initial task, the identical part of the instructions informed subjects of the nature of the task they were about to undertake. Additional to this, the control task group given the person interaction task first were told "Your aim is to shift Clegg to the '*Polite*' level and maintain him at that level". The control task group given the factory task first was told "Your aim is to make the factory produce an output of *6000 tons* and keep it at that level". The pattern search group given the person interaction task first was told "Your aim is to establish under what pattern Clegg is reacting". The pattern search group given the factory task first was told "Your aim is to determine the pattern that is being followed that leads the factory to produce the various levels of sugar outputs". To remind subjects of their respective goals, the goal of their task was permanently displayed on a piece of paper attached to the bottom of the screen. The rest of the experiment was identical for each group.

On each trial for the person interaction task, Clegg's and the subject's responses were displayed on the screen. In the case of the factory task, the previous factory output and the subject's new work force size were displayed. The information scrolled up the screen so that it was possible to see the previous six trials on the screen at any one time. Clegg's responses of , *Indifferent, Polite,* or *Very Polite* (*5000, 6000* or *7000 tons* for the factory task) were scored as correct to allow for the random element in the equation that produces Clegg's response (or the factory's output). The equation that controlled the output from the sugar production factory was identical to that controlling Clegg's output. In the case of the factory task, the numbers from 1 to 12 were associated with the size of work force or size of factory output (e.g. *100 employees* = 1, *1000 tons* = 1). Following the learning phase the instructions for the test phase and the test phase itself were presented.

The test phase was identical to the learning phase for the control task group except that the goal was different. The new goal for the person interaction task for both the control task subjects and pattern search subjects was to make Clegg be *Very Friendly* and maintain him at that level. The new goal for the factory task was *9000 tons* (the equivalent of *Very Friendly*). Clegg's response of *Friendly, Very Friendly* or *Affectionate* (*8000, 9000* or *10000 tons* for the factory task) were scored as being on target to allow for the random element in the equation. Following the test phase the questionnaire was presented to the subjects. They had no warning that this was coming.

The Questionnaire: This consisted of prediction questions and rule description questions identical to that described in the Method of Experiment 1. Subjects in both goal groups who learned the factory task had a comparable set of prediction questions.

Following the questionnaire the subjects entered the transfer phase. This was identical to the test phase except that subjects tried to achieve the equivalent goal on the new task. Both goal groups that learned the person interaction task first were given instructions explaining the nature of the factory task and were set a goal of making and maintaining sugar production at *9000 tons* output. The subjects that learned the factory task first were given instructions explaining the nature of the person interaction task and were set a goal of making and maintaining Clegg's output at the *Very Friendly* level.

Finally subjects were presented with 3 questions that asked subjects to rate the similarity of the two task.

1) One question asked for ratings of <u>overall similarity</u>: "How did you rate the overall similarity between the factory task and the person interaction task?"
2) A second question asked for ratings of <u>underlying similarity</u>. This was the same as question (1) plus, "...., ignoring differences occurring from the appearance of the two tasks on the screen?"
3) A third question asked for ratings of <u>strategic similarity</u>. This was the same as question (1) plus, "...., with respect to the optimum strategy needed to reach target levels?"

Subjects were told to rate similarity on a scale of (1) extremely different to (5) extremely similar. These questions were the same as those used by Squire & Frambach (1990).

Procedure: Subjects were randomly allocated to one of the two groups then to one of the two conditions within each group. As mentioned above, the groups received identical initial instructions apart from one sentence. This sentence dictated the aim of that particular group for their learning phase of the experiment. Apart from this, relevant to the order of their tasks, the experiment was identical for both groups.

The instructions explaining the nature and aim of the subjects' initial learning task were presented first. These were followed by the learning phase of the experiment. On completion of this phase subjects from the four groups received instructions describing their new aim for the test phase and then the test phase started. Following the test phase the subjects were presented with instructions for the prediction part of the questionnaire. After completing the prediction section subjects were given a pen and paper and were asked to answer the two rule description questions. Subjects then were given instructions for their new and final task (the transfer phase) and then it began. Clegg initiated learning, test and transfer phases by displaying one of the three adjectives centred on *Polite* (or in the case of the factory task, an output centred on *6000 tons*). Finally, subjects were presented on screen with the 3 similarity perception questions. Their answers were entered into and recorded by the computer.

Throughout the experiment, all instructions appeared on the computer screen but were also read out to the subjects. The experimenter stayed with the subject throughout the experiment in order to answer any arising questions.

RESULTS

Performance during learning and testing

As mentioned before, during the learning phase of the experiment trials were scored as correct for the control task group if they got a response from Clegg of *Indifferent, Polite* or *Very Polite* if they did the person interaction task first or *5000, 6000* or *7000 tons* if they did the factory task first. This takes into account the random element of the underlying equation of the two tasks. Due to the lack of specific aim for the pattern search group during the learning phase, no measure could be made for its performance during the first set of trials. For all groups, trials were scored as correct during the testing and transfer phases if subjects got a response from Clegg of *Friendly, Very Friendly* or *Affectionate* or an output from the factory of *8000, 9000* or *10000 tons.*

The total number of correct trials during each phase of the experiment and the total number of correct trials for each half of an experimental phase were recorded. The mean numbers of correct trials were calculated for each set of subjects in each group in each of these categories.

Learning Trials

The scores for the learning phase can be seen in Figure 6.1.

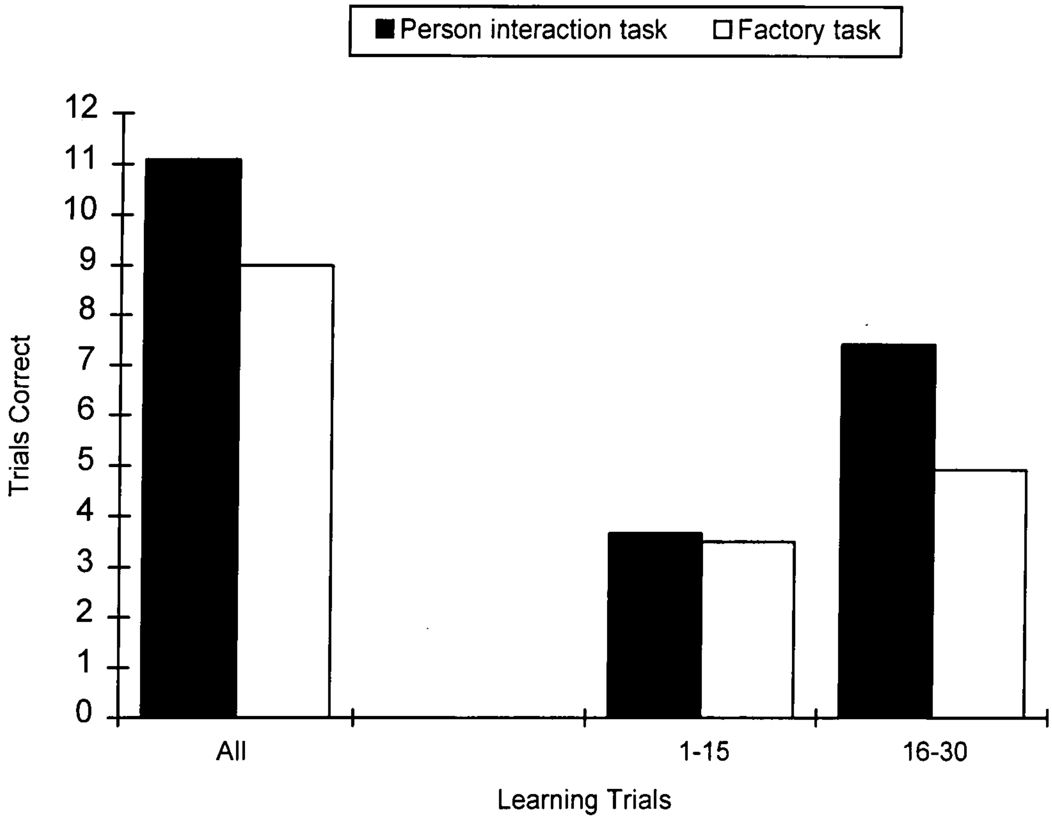


Figure 6.1: Mean number of correct trials in the learning phase for the control task subjects. Data are shown for all 30 trials combined and for the first and second 15 trials.

The data in Figure 6.1 were analysed using a 2 (task) by 2 (trial block) analysis of variance with repeated measures on the last factor. The main effect of task just failed to reach significance, $F(1,22) = 3.61$, $p = 0.071$: there was a tendency for learning to be better on the person interaction task than the factory task. The main effect of trial block was significant, $F(1,22) = 6.17$, $p = 0.021$: there were more correct trials in the second 15 than in the first. The interaction between trial block and task was not significant. (See Appendix 6 for the ANOVA tables and full sets of t-tests for this experiment, pg. 251).

The Test Trials

The scores for the test phase can be seen in Figure 6.2.

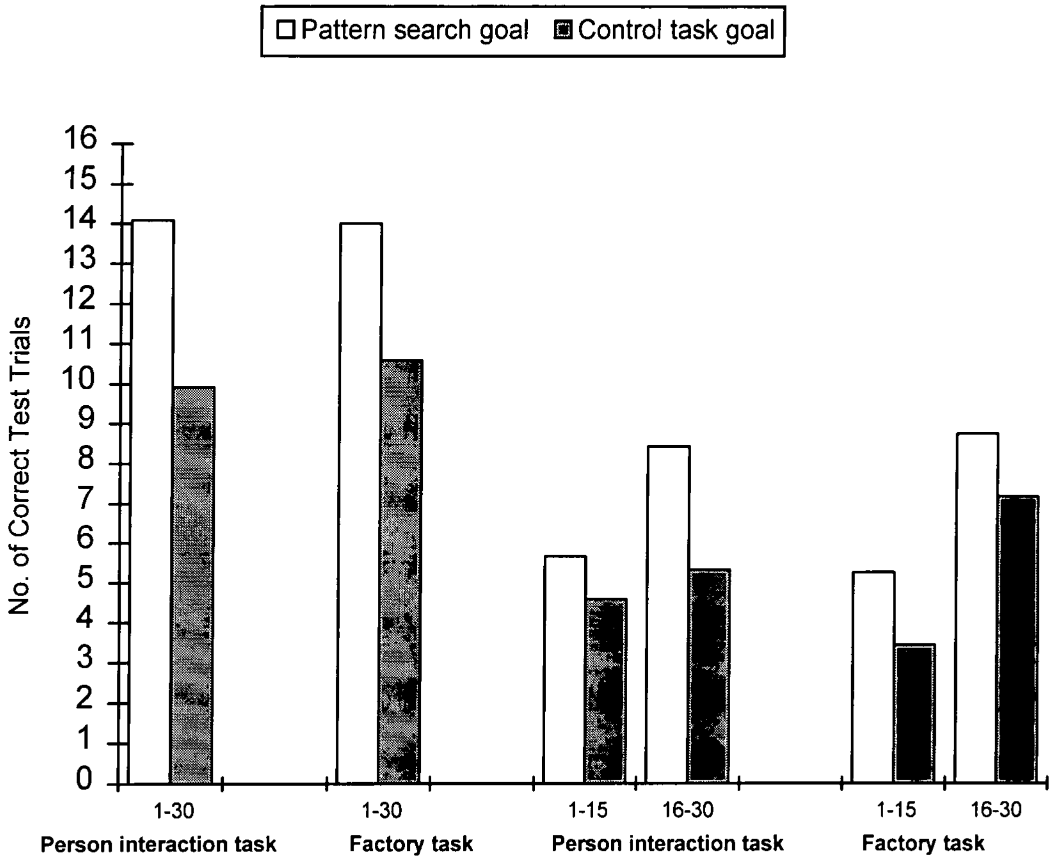


Figure 6.2: Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.

The data in Figure 6.2 were analysed using a 2 (task) by 2 (learning goal) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(1,44) = 7.16$, $p = 0.01$: subjects with a pattern search goal performed better during the test phase than those with a control task goal. There was also a main effect of trial block, $F(1,44) = 27.06$, $p < 0.001$: there were more correct trials in the second 15 than in the first. However the interaction between trial block and task just failed to reach significance, $F(1,44) = 3.29$, $p = 0.076$: Inspection of Figure 6.2, shows there was a tendency for there to be more improvement during the test phase for the factory task. There were no other significant effects or interactions.

Transfer Of Learning Between Learning And Test Phases

Comparing learning and test phase performances should throw some light on how the subjects in the control task group managed when their task was still the same, but the goal was switched to a new target. Two comparisons were made to examine this: (i) A comparison between the number of correct trials during the last *half* of the learning phase and the first *half* of the test phase and, (ii) A comparison between the *total* number of correct trials during the learning and test phases. For comparison (i) a 2 (task) by 2 (trial block) mixed analysis of variance revealed a main effect of task, $F(1,22) = 6.23$, $p = 0.021$: performance was better for the person interaction task overall on the last half of the learning phase and the first half of the test phase. There was also a main effect of trial block, $F(1,22) = 6.28$, $p = 0.02$: there were more trials correct in the last half of the learning phase than in the first half of the test phase. The interaction between the two effects was not significant. So, initially subjects' performance dropped when the control task goal changed. For comparison (ii) a 2 (task) by 2 (phase score) mixed analysis of variance revealed no main effect of task or of phase score. The interaction between the two just failed to reach significance, $F(1,22) = 3.17$, $p = 0.089$: there was a tendency for the factory task to perform better during the test phase than during the learning phase.

The Prediction Questions

Answers to the New, Old-correct, and Old-wrong situations were scored in the same way as in Experiment 1 (see pg. 28). As in the Experiment 1, responses were discarded (7.5% of the data) if a selected trial type (e.g. Old-wrong) had been responded to differently on a second occasion (e.g. making it also an Old-correct trial type). The resulting mean percentage of correct responses are shown in Figure 6.3. Due to some of the data being discarded, these results are shown as percentages.

The data in Figure 6.3 are analysed using a 2 (learning goal) by 2 (task) by 3 (question type) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(1,44) = 41.24$, $p < 0.001$: subjects with a pattern search goal performed better than those with a control task goal overall on the prediction questions. There was also a main effect of question type, $F(2,88) = 4.85$, $p = 0.01$, and as predicted an interaction of this main effect with learning goal, $F(1,88) = 7.42$, $p = 0.001$. There were no other significant effects or interactions.

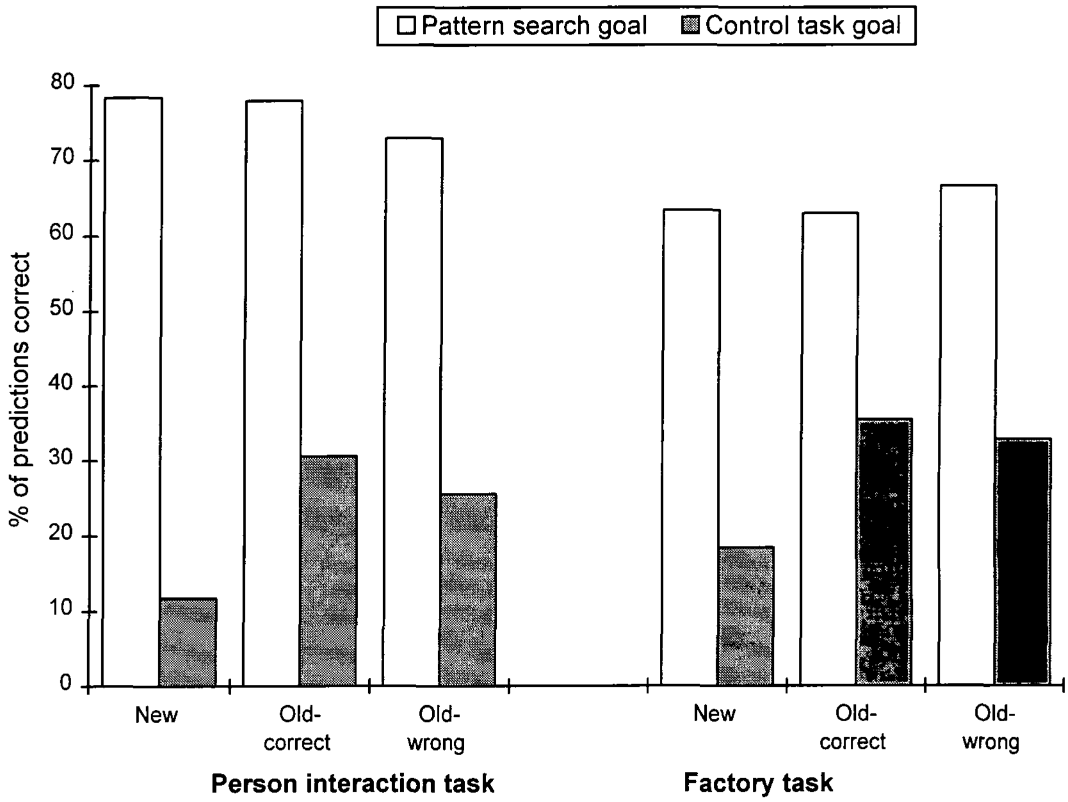


Figure 6.3: Mean percentage of correct responses to each category of prediction question for each group. Total prediction question scores are in the top graph, and individual types below.

To explore the interaction between question type and learning goal, within group comparisons were carried out. For subjects with a control task goal, as predicted, Wilcoxon matched paired tests showed a significantly higher score for Old-wrong questions compared to New questions, $Z = -3.62$, $p < 0.001$, and a significantly higher score for Old-correct compared to New questions, $Z = -2.94$, $p = 0.003$. There was no significant difference between Old-wrong and Old-correct questions. Also as predicted, for subjects with a pattern search goal there was no significant differences between any of the question types.

As in Experiment 1 the results were also analysed defining the prediction situations by the last two elements (again discarding questions for which the situations could have occurred with both a correct and an incorrect response) - see page 31 for a longer explanation of why the results were reanalysed in this way. The results can be found in Appendix 6 on pg. 252. These reanalyses showed an identical pattern of statistics with all the results that were significant before remaining significant (all p values < 0.05).

Correlations between control performance and predictions: To examine how much the predictions relied upon control performance, total prediction scores were correlated with the number of correct trials during the test phase. For the pattern search group, Spearman rank correlation coefficients were 0.81 for the set of subjects doing the person interaction task and 0.87 for those doing the factory task. Both these coefficients were significant ($p < 0.002$). This suggests that irrespective of task, for subjects with a pattern search goal, predictions were related to control performance. For the control task group, Spearman rank correlation coefficients were -0.27 for those subjects doing the person interaction task, and 0.31 for those doing the factory task. Neither of these correlations was significant ($p > 0.2$). For the control task group total prediction questionnaire scores were also correlated with the number of correct trials during the learning phase; -0.003 for those doing the person interaction task, and 0.28 for those doing the factory task. Again neither of these correlations was significant ($p > 0.2$).

The Rule Description Questions

Subjects' answers to the two rule description questions were ranked in the same way as in Experiment 1 (see pg. 32). Both judges categorised the answers identically. These categorisations can be seen in Table 6.1.

Fisher exact probability tests compared the number of answers in the *No information or wrong* category and in the *Correct* category. There was no significant difference between the two sets of subjects in the pattern search group or between the two in the control task group. For the person interaction task, there was a difference between the subjects with a pattern search goal and those with a control task goal ($p < 0.0001$). The same applied for the factory task ($p < 0.0001$). For both these tasks, the sets of subjects with the pattern search goal were getting more answers in the *correct* category and less answers in the *No information or wrong* category than those subjects with the control task goal. Thus, subjects with a pattern search goal were better than those with a control task goal at producing answers that contained declarative knowledge. In the stricter test of explicit learning, as used in Experiment 1 (see pg. 33), the number of questions in the partially correct and correct categories were added together. Fisher exact probability tests showed an identical pattern to the above one (all $p < 0.002$).

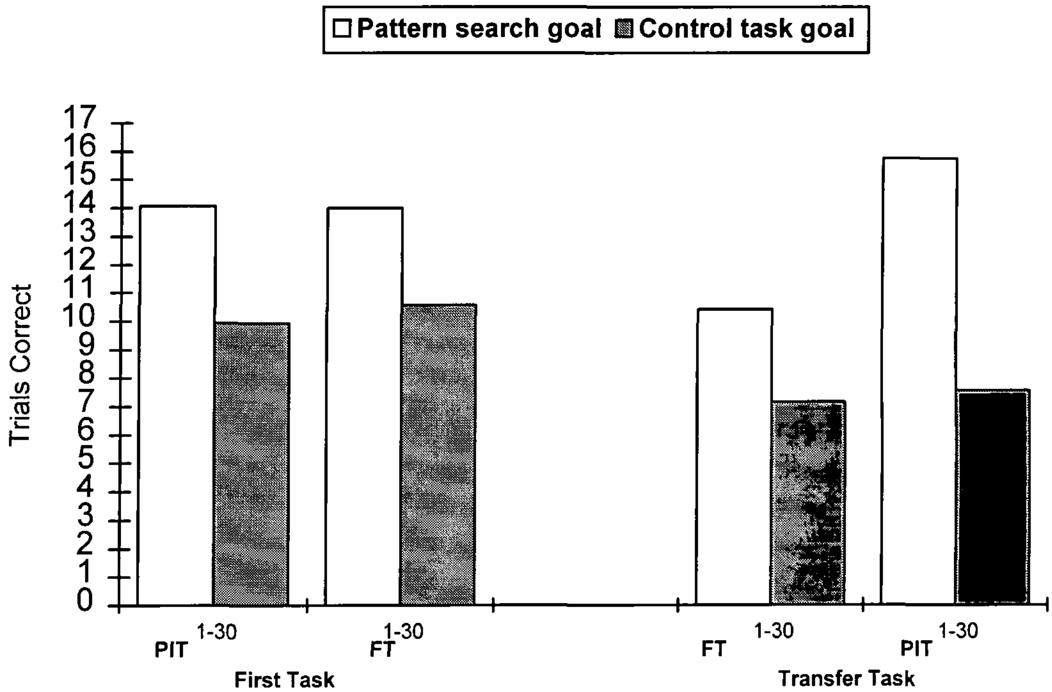
Table 6.1
Ranking of answers to the rule description questions into each category for the two groups.

	Category (PIT) = person interaction task (FT) = factory task	No information or Wrong	Partially Correct	Correct
Task	**Group**			
PIT	control task	**9**	**3**	**0**
PIT	pattern search	**1**	**3**	**8**
FT	control task	**8**	**4**	**0**
FT	pattern search	**0**	**2**	**10**

Between Task Transfer

The mean number of correct trials for each set of subjects in each group, for the entire transfer phase and each half of the transfer phase can be seen in Figure 6.4 below.

To explore how subjects coped when their task changed a 2 (task order: person interaction → factory vs factory →person interaction) by 2 (learning goal) by 2 (phase score: test vs transfer) analysis of variance with repeated measures on the last factor revealed a main effect of learning goal, $F(1,44) = 14.60$, $p < 0.001$: overall performance during the test phase and the transfer phase was better for subjects with a pattern search goal. There was also a main effect of phase score, $F(1,44) = 9.78$, $p = 0.003$: subjects performed better during the test phase than during the transfer phase. However, there was an interaction between phase score and task order $F(1,44) = 4.44$, $p = 0.041$, and also a three way interaction between phase score, learning goal and task order, $F(1,44) = 5.34$, $p = 0.026$.



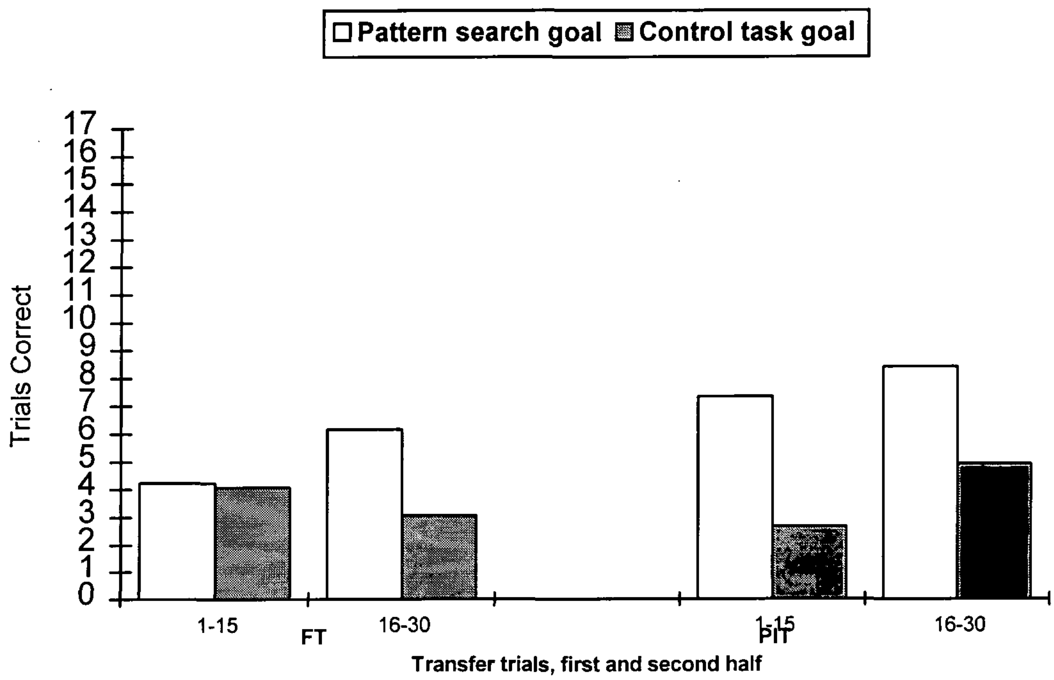


Figure 6.4: Mean number of correct trials in the transfer phase (and test phase for total scores) for each group. Data are shown for all 30 trials combined in the top graph and for the first and second 15 trials in the bottom. PIT = person interaction task, FT = factory task.

Within group comparisons looking at phase score showed that performance on the transfer phase scores was significantly worse than that on the test phase for PIT→FT control task group, $t(11) = -2.83$, $p < .02$, FT→PIT control task group, $t(11) = -7.54$, $p < .001$ and for the PIT→FT pattern search group, $t(11) = -2.48$, $p < .04$ The only group to show no drop in performance (and therefore the only group to show complete transfer of learning) was the pattern search group changing from the *factory task* to the *person interaction task*.

To strengthen the claim of transfer occurring for this latter group. It was predicted that overall on the transfer phase, the transferring group should outperform the groups not demonstrating transfer. With this aim in mind the four sets of subjects were compared to each other. The F and *p* values can be seen in Table 6.2 below. For all non-significant results *p* values are > 0.1. As predicted the FT→PIT pattern search group outperformed the other three groups (see row (4), columns (1), (2) & (3)). For the other significant results in Table 6.2, the PIT→FT pattern search group is performing better than the control task subjects.

Table 6.2
Between groups comparisons for the total transfer phase score

Label	Task order	learning goal	(1)	(2)	(3)
(1)	PIT→FT	control task group			
(2)	PIT→FT	pattern search group	5.94 0.023		
(3)	FT→PIT	control task group	0.23 ns	4.77 0.04	
(4)	FT→PIT	pattern search group	13.03 0.002	4.32 0.050	12.0 0.002

F(1,22) value PIT = person interaction task ns = not
p value FT = factory task significant

Correlation of transfer phase with earlier phases: To explore how much relationship there was between performance on the novel task and performance on the earlier task, the total transfer phase scores were correlated with the total test phase scores, and for the two sets of subjects in the control task group, also with the total learning phase scores. For the pattern search group, Spearman Rank correlation coefficients were, 0.63 ($p < 0.04$) for the subjects that switched from person interaction to factory task and, 0.77 ($p < 0.005$) for the subjects that switched from factory to person interaction. These significant correlations suggest that for both sets of subjects with a pattern search goal, their system for controlling the initial task was closely related to that used in the novel task. For the control task subjects switching from person interaction to factory task, the coefficient with the test phase was 0.004 ($p > 0.9$) and with the learning phase was -0.28 ($p > 0.3$). For the control task group switching from factory to person interaction task, the coefficient with the test phase was -0.14 ($p > 0.6$) and with the learning phase was 0.39 ($p > 0.2$). None of these coefficients are significant indicating that for the subjects with a specific learning goal, performance in the first task was not related to performance in the second task.

Opinions On The Similarity Of The Two Tasks

The average similarity rating for each type of similarity probed in the final 3 questions can be seen in Table 6.3 (the scale runs from 1 = extremely different up to 5 = extremely similar). The data in Table 6.3 were analysed using a 2 (learning goal) by 2 (task order) by 3 (type of similarity: Overall vs Underlying vs Strategic) analysis of variance with repeated measures on the last factor. The results revealed a significant main effect of learning goal, $F(1,44) = 12.71$, $p = 0.001$: on average for the three similarity types, subjects with a pattern search goal gave the two tasks a higher rating of similarity than subjects with a control task goal. There was also a main effect of type of similarity, $F(2,88) = 17.41$, $p < 0.001$ and a significant interaction between type of similarity and learning goal $F(2,88) = 4.02$, $p = 0.021$. There were no other significant interactions. To explore the interaction between type of similarity and learning goal the data were collapsed over task order and within groups comparisons were made between the different types of similarity questions. Subjects with a control task goal gave a comparable ratings for all three similarity types. Subjects with a pattern search goal gave a comparable rating of similarity for Underlying Similarity and Strategic Similarity and a higher rating for both these types compared to Overall Similarity. For Underlying Similarity vs Overall Similarity, $Z = -3.82$, $p < 0.001$ and for Strategic Similarity vs Overall Similarity, $Z = -3.62$, $p < 0.001$.

Table 6.3
Similarity ratings of the two tasks for each set of subjects

	Category	Overall Similarity	Underlying Similarity	Strategic Similarity
PIT = person interaction task FT = factory task				
Task order	**Learning goal**			
PIT→FT	control task	**2.00**	**2.33**	**2.33**
PIT→FT	pattern search	**2.75**	**3.50**	**3.50**
FT→PIT	control task	**2.33**	**2.75**	**2.67**
FT→PIT	pattern search	**2.92**	**4.00**	**4.25**

DISCUSSION

How did the groups learn? It is clear that learning goal (control task or pattern search) determined the type of learning, irrespective of the task (person interaction or factory). Subjects with a control task goal learnt differently from those with a pattern search goal. Subjects with a control task goal learnt both tasks in a similar fashion despite the fact that the tasks were different and this was also true for the subjects with a pattern search goal.

The evidence indicates that the pattern search goal subjects learnt through explicit hypothesis testing, while, as in previous studies (e.g. Berry & Broadbent, 1984; Dienes & Fahey, 1995, Geddes & Stevenson, in press), control task goal subjects learnt through instance learning. It is deduced that the pattern search subjects learnt explicit rules from the following points: (1) The similarity of performance on all prediction question types, particularly, being able to do just as well at New questions as at the Old-wrong and Old-correct questions; (2) The success at answering the rule description questions; (3) The significant positive correlations between total questionnaire scores and test phase trial performance. Evidence for the control task group learning instances is inferred from the following points: (1) The significant difference in performance on the different prediction question types, namely, poor performance on the New questions; (2) The poor answers to the rule description questions; (3) The lack of a significant positive correlation between total questionnaire score and trial performance. The notion that the two types of group learned differently is further reinforced by the comparisons between the groups; the pattern search groups performed better on both the prediction questions and the rule description questions, indicating that they had acquired significantly more verbalisable knowledge than the control task group. These results support the hypothesis that groups with a non control task goal should learn through explicit hypothesis testing while groups with a control task goal should learn through implicit instance learning. Importantly, these results clearly indicate that this hypothesis holds true for the previously untested factory task as well as for the person interaction task. Support of this hypothesis for the person interaction task replicates the earlier findings reported in the rest of the thesis. Having established that the two groups of subjects learned differently, the differences in transfer of learning between rule learners and instance learners can now be explored.

How did learning transfer? Complete transfer of learning occurred for one set of subjects only - the pattern search goal subjects whose first task was the factory task. There is some evidence that the other pattern search subjects had some link between the first task and the second, but proper transfer of learning did not occur. No transfer of learning occurred for the control task subjects.

It is deduced that learning on the factory task was transferred to the person interaction task for the pattern search goal subjects from two points. (1) The comparable performances between test and transfer phase. This obviously indicated that there was no drop in control ability between the two tasks. It would be safe to assume that for this group, what was learnt in the factory task enabled subjects to then control the person interaction task at a similar level of competence. (2) The significant positive correlation between test phase and transfer phase. This suggests a close link between performance in the test phase and performance in the learning phase. It is likely that control methods used by subjects in the test phase were then transferred to the transfer phase. It is deduced that learning on the person interaction task did not fully transfer to the factory task for the pattern search group from the following points. (1) The significantly lower level of performance during the transfer phase when compared to the test phase. This showed an overall drop in performance, clearly indicating that what was learnt during the person interaction task did not wholly transfer to the factory task. (2) The positive correlation between test phase and transfer phase performances however, suggests that there were some similarities between methods used on the two tasks. For the control task subjects it is deduced that learning on the first task did not transfer at all to the second task (which ever one it might have been) from the following two points. (1) The significant lower levels of performance during the transfer phase when compared to the test phase. (2) The lack of significant correlation between transfer phase performance and performance in either the learning or the test phases. This indicates that performance in the first task was not related to performance in the other.

In summary, the transfer of learning results describe a pattern of (a) transfer of learning occurring for the pattern search group switching from factory task to person interaction task, b) learning not transferring properly for the pattern search goal where the switch in task is in the opposite direction, however, there still being some connection between performance on the two tasks, (c) no transfer of learning occurring for control task subjects. Further support for this pattern comes from examining the between groups comparisons of performance during the transfer phase; The subjects in the pattern search group who started on the factory task outperformed the other three groups. The pattern search group that started on the person interaction task outperformed the two sets of subjects in the control task group, while these two sets of subjects performed comparably.

How did the groups recognise the underlying similarity of the two tasks? Those subjects learning instances were less likely to label the two tasks as similar than those subjects learning rules. This can be deduced from the fact that there was an overall effect of learning goal on similarity ratings: subjects with a pattern search goal rated the two tasks as more similar than subjects with a control task goal. However, rule learners strength over instance learners at recognising the similarity of the two tasks comes from recognising the underlying and strategic similarities. This is deduced from the fact that within group comparisons showed that while the instance learners performed equally on all three measures of similarity, pattern search subjects gave significantly higher ratings of similarity for the second two questions (those measuring underlying and strategic similarity). These points can be taken to add weight to a number of propositions voiced so far. It supports the idea that subjects given a pattern search goal were learning explicit rules and those with a control task goal were learning instances. Rule learners, people with more awareness of the underlying patterns of the tasks, should be more likely to explicitly recognise underlying and strategic similarities of the tasks than instance learners, people with little detailed awareness of the underlying pattern of the two tasks. It supports the idea that transfer of learning was occurring for the pattern search subjects that started with the factory task. If these subjects were learning through explicit hypothesis testing and rule deduction and were using control methods from the factory task on the person interaction task it would be expected that they would recognise the underlying similarity of the two tasks. The results of the similarity ratings support the idea that for the pattern search subjects who started with the person interaction task, transfer of learning can't be completely ruled out. The realisation by these subjects that the two tasks had underlying similarities would explain why there was a significant positive correlation between test and transfer phases. It suggests that at some stage during the new task, subjects realised the similarity and were able to work out how to act as they did for the old task on the new task. Exactly why both sets of subjects in the pattern search group did not exhibit complete transfer of learning shall now be considered.

Considering the recognition of the underlying similarities of the two tasks for the pattern search group it is initially surprising that complete transfer of learning was not shown. Possibly the similarity was recognised later on in the task than for the successful pattern search subjects, and thus what was learnt in the person interaction task could not be transferred to the factory task till this late recognition was made. Another possibility is that the pattern search subjects had trouble applying their initial learning on the person interaction task to the factory task because the latter was more complicated than the person interaction task. This is considered below. Firstly though whether the factory task was simply harder than the person interaction task is considered. Stanley et al (1989) provided evidence for this for groups given a control task goal. The results in this experiment however, while not clearly contradicting these previous findings, suggest that subjects performance on the two tasks was more similar to each other than was found by Stanley et al. All the subjects in the pattern search group performed comparably overall for the test phase and overall for the questionnaire. Comparisons for the control task group showed only a marginally significant difference for the learning phase. For the prediction questions, control task subjects performed comparably irrespective of task. Despite the findings giving no strong suggestion that the factory task is harder than the person interaction task there are some key differences that could explain why complete transfer of learning was not seen from the person interaction task to the factory task.

The person interaction task has inputs and outputs coming from an identical pool of possible options - the twelve behaviours. This means that inputs and outputs are in the same currency so to speak. Subjects attempting to establish the pattern that the system works under simply need to work out how new outputs are related to the previous output and input. They can explore this by using the underlying scale that both the inputs and outputs fall on - the behaviours from *Very Rude* to *Loving*. The factory task is not as simple. Inputs and outputs are in different currencies, with input being in work force size and output being in sugar tonnage. Apart from the 'currency' being different the possible numerical values within each 'currency' are different by a factor of ten. 100 workers is equivalent to 1000 tonnes of sugar on the underlying scale on which the pattern is based. Before subjects can attempt to establish any pattern they have to equate sugar tonnage with work force size and equate 100 workers with 1000 tonnes of sugar. Considering this, it is surprising that the study did not show more substantial differences in performance between the factory and person interaction task. This extra equating element of the factory task could be responsible for the lack of evidence for transfer of learning from the person interaction task to the factory task for the pattern search subjects. In the second experiment reported in this chapter this possibility is tested.

For the second experiment the factory task was simplified so that the equating element of the task is no longer necessary. To do this the inputs and outputs were simply made the same 'currency' and the cover story modified to take account of this. So, the second experiment is identical to the first except that the factory task has been modified. The main hypothesis is that transfer of learning should now also occur for the pattern search subjects switching task from person interaction to factory task i.e. learning on the person interaction task should now transfer to the simplified factory task as well as vice versa.

Experiment 6b - METHOD

Subjects: The 48 volunteer subjects were Durham University graduate and undergraduate students, aged between 18 and 24.

Design: Exactly as in the previous experiment, there were two main groups of subjects - a control task group and a pattern search group. Within each group, half the subjects had one task first (the person interaction task) while the other group had the other task first (for this experiment it was the simplified version of the sugar production task). As in Experiment 6a, subjects were randomly allocated to one of the four sets of subjects. The design was identical in every way to the previous experiment. The only variation being that the factory task was altered to make it simpler.

The simplified version of the sugar factory production task : This version of the task is identical to the previous version except that the inputs and outputs are slightly different and the cover story varies to take account of this. Subjects were asked to imagine that they were in charge of a sugar production factory in an underdeveloped country. They could control the rate of sugar production simply by entering how many thousands of tonnes of sugar they wanted the factory to produce (with a possible range of 1000 to 12000 tonnes). The market place would require a certain tonnage of sugar reported to the nearest 1000 tonnes (again this could be anything from 1000 to 12000 tonnes). To tell the computer what size of output they wanted to set the factory to produce, subjects simply had to type in a number from 1 to 12 representing the number of thousands of tonnes of sugar. Subjects were told that as the output from the factory varied, the market requirement would also vary. The market place would start off requiring a certain tonnage of sugar. Once subjects had been told how much sugar the market place required, they would then enter in the next size of output they wanted the factory to produce. Once subjects had set the new output the resulting requirement from the market place would be displayed. Then subjects again altered the size of the output, and so on. The market place requirement was calculated by the same equation that predicted Clegg's next response, and that controlled the factory in the previous experiment (see the Method from Experiment 6a).

Procedure: The procedure was identical to that of the previous experiment.

RESULTS

The Learning Trials

The scores for the learning phase can be seen in Figure 6.5. The data in Figure 6.5 were analysed using a 2 (task) by 2 (trial block) analysis of variance with repeated measures on the last factor. The main effect of trial block was significant, $F(1,22) = 8.63$, $p = 0.008$: there were more correct trials in the second 15 than in the first. There were no other significant effects or interactions. (See Appendix 6 for the ANOVA tables and full sets of t-tests for this experiment, pg. 255).

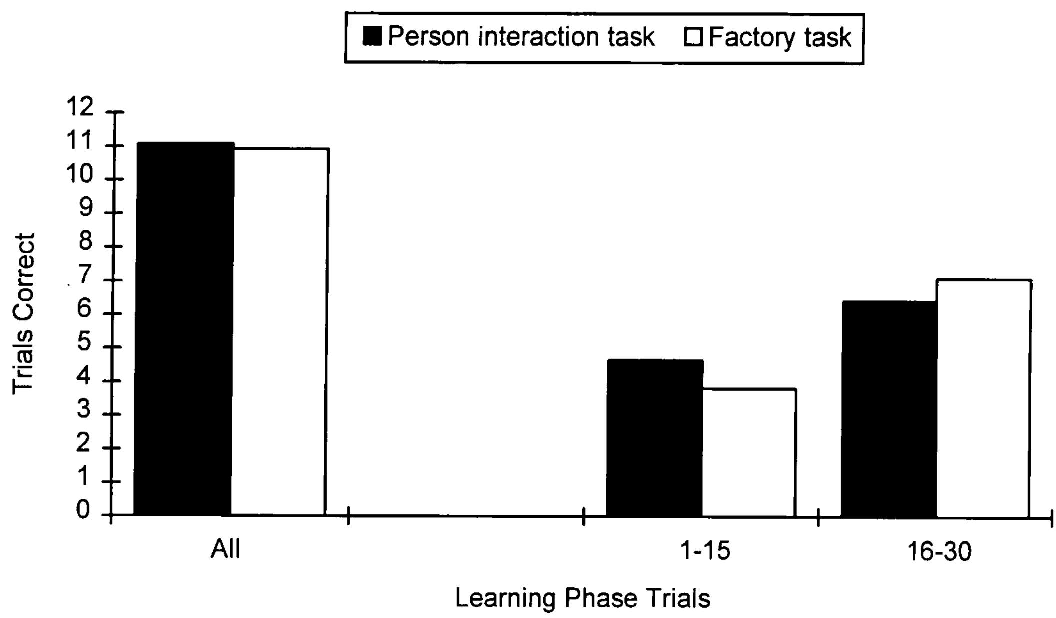


Figure 6.5: Mean number of correct trials in the learning phase for the control task subjects. Data are shown for all 30 trials combined and for the first and second 15 trials.

The Test Trials

The mean number of correct trials for each set of subjects within each group, for the entire test phase and each half of the test phase can be seen in Figure 6.6.

The data in Figure 6.6 were analysed using a 2 (task) by 2 (learning goal) by 2 (trial block) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(1,44) = 7.56$, $p = 0.009$: subjects with a pattern search goal performed better during the test phase than those with a control task goal. There was also a main effect of trial block, $F(1,44) = 66.59$, $p < 0.001$: there were more correct trials in the second 15 than in the first. There were no other significant effects or interactions.

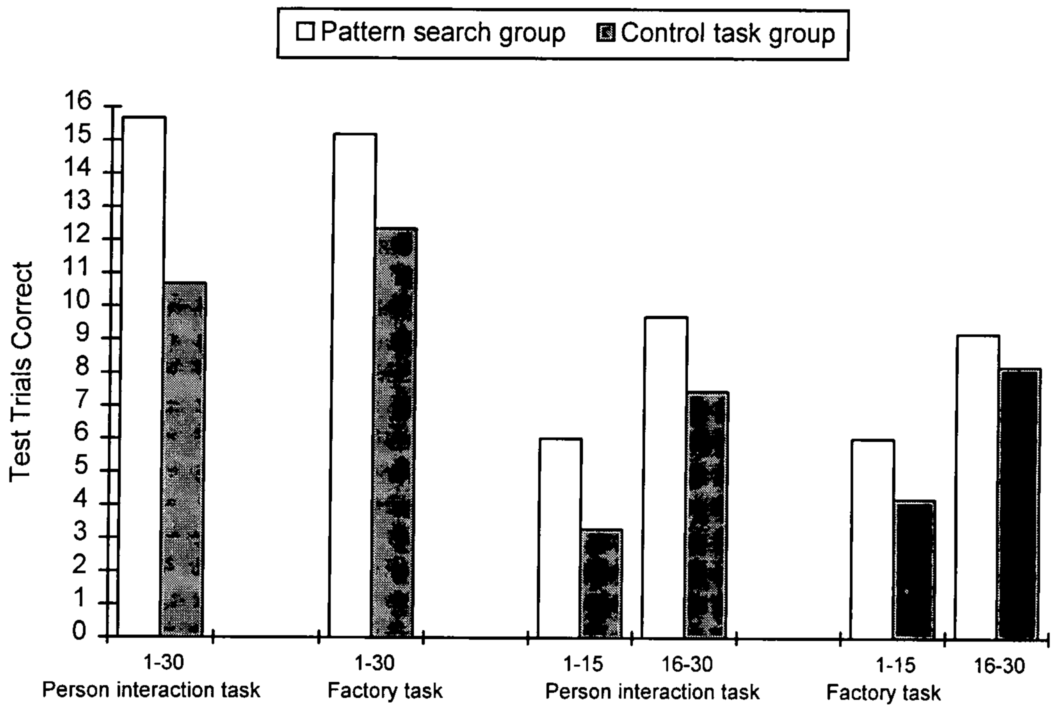


Figure 6.6: Mean number of correct trials in the test phase for each group. Data are shown for all 30 trials combined and for the first and second 15 trials.

Transfer Of Learning Between Learning And Test Phases

Two comparisons were made to examine transfer to a new specific goal by the control task subjects: (i) A comparison between the number of correct trials during the last *half* of the learning phase and the first *half* of the test phase and, (ii) A comparison between the *total* number of correct trials during the learning and test phases. For comparison (i) a 2 (task) by 2 (trial block: last half learning phase vs first half test phase) mixed analysis of variance showed a main effect of trial block, $F(1,22) = 21.56$, $p < 0.001$: there were more trials correct in the last half of the learning phase than in the first half of the test phase. There was no main effect of task or significant interaction. Therefore, initially subjects performance dropped when the control task goal changed. For comparison (ii) a 2 (task) by 2 (phase score: learning vs test) mixed analysis of variance revealed no main effects and no significant interactions. Irrespective of task, subjects performance between learning and test phases and overall during learning and test phases was comparable.

The Prediction Questions

All subjects provided enough correct and incorrect questions on the test trials for the prediction questions. The same problem with question generation explained in Experiment 6a (see pg. 169) applied here and problematic data were removed (7.5% of the data). Therefore, again, results are reported in percentages. Due to some of this data being removed, one of the control task subjects starting on the factory task had no questions for the Old-correct question type. The mean percentages of correct responses to each question type are shown in Figure 6.7. The data in Figure 6.7 are analysed using a 2 (learning goal) by 2 (task) by 3 (question type) analysis of variance with repeated measures on the last factor. The results revealed a main effect of learning goal, $F(1,43) = 21.35$, $p < 0.001$: subjects with a pattern search goal performed better than those with a control task goal overall on the prediction questions. There was also a main effect of question type, $F(2,86) = 5.11$, $p = 0.008$, and as expected an interaction of this main effect with learning goal, $F(2,86) = 4.68$, $p = 0.012$. There were no other significant effects or interactions.

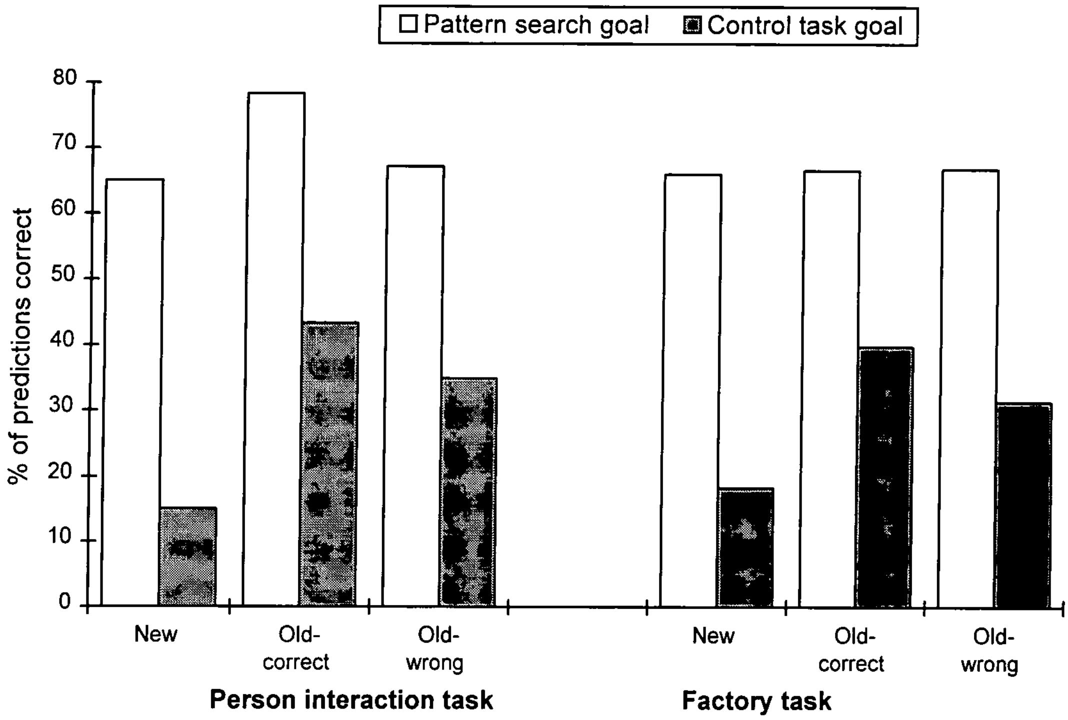


Figure 6.7: Mean percentage of correct responses to each category of prediction questions for each group.

To explore how the different learning goals affected performance across questions type data were collapsed over task and within group comparisons were made between the different question types. For subjects with a control task goal, Wilcoxon matched paired tests showed a significant difference between Old-wrong and New scores ($Z = -3.12$, $p = 0.002$), and between Old-correct and New scores ($Z = -3.46$, $p < 0.001$). However, no significant difference was found between scores for Old-wrong and Old-correct questions. For subjects with a pattern search goal there was no significant differences between any of the question types.

As in Experiment 1 the results were also analysed defining the prediction situations by the last two elements (again discarding questions for which the situations could have occurred with both a correct and an incorrect response) - see page 31 for a longer explanation of why the results were reanalysed in this way. The results can be found in Appendix 6 on pg. 256. These reanalyses showed an identical pattern of statistics with all the results that were significant before remaining significant (all p values < 0.05).

Correlations between control performance and predictions: For the pattern search group, Spearman rank correlation coefficients were 0.74 for the set of subjects doing the person interaction task and 0.81 for those doing the simplified factory task. Both these coefficients were significant ($p < 0.007$). As in Experiment 6a, this suggests that irrespective of task, for subjects with a pattern search goal, control performance and prediction performance were related. For the control task group, Spearman rank correlation coefficients were -0.06 for those subjects doing the person interaction task, and 0.27 for those doing the simplified factory task. Both these correlations were not significant ($p > 0.3$). For the control task group total prediction scores were also correlated with the number of correct trials during the learning phase; -0.23 for those doing the person interaction task, and -0.11 for those doing the simplified factory task. Again both these correlations were not significant ($p > 0.5$).

The Rule Description Questions

As in Experiment 6a, subjects' answers to the two questions (asking how to control Clegg and asking under what pattern Clegg was operating) were treated together as subjects generally answered only one of the questions and included information in that answer that was relevant to both questions. The answers were judged by two judges and ranked into the three categories; *No information or Wrong, Partially Correct, Correct.* Both judges ranked the answers identically. These rankings can be seen in Table 6.4.

Table 6.4
Ranking of answers to the rule description questions into each category for the two groups.

(PIT) = person interaction task (FT) = factory task	**Category**	No information or Wrong	Partially Correct	Correct
Task	**Group**			
PIT	control task	**11**	**1**	**0**
PIT	pattern search	**0**	**1**	**11**
FT	control task	**8**	**2**	**2**
FT	pattern search	**0**	**1**	**11**

Fisher exact probability tests compared the number of answers in the *No information or wrong* category to those in the *Correct* category. There was no significant difference between the two sets of subjects in the pattern search group or between the two in the control task group.

For the person interaction task, there was a difference between the subjects with a pattern search goal and those with a control task goal ($p < 0.001$). The same applied for the factory task ($p < 0.001$). For both these cases subjects with the pattern search goal were getting more answers in the *correct* category and less answers in the *No information or wrong* category than those subjects with the control task goal. Thus, subjects with a pattern search goal were better than those with a control task goal at producing answers that contained declarative knowledge. It might be argued however, that too strict a criterion was used to categorise *Correct* and that with a looser criterion pairs of groups may have been more similar. Adopting a more stringent criteria for lack of explicit knowledge, more Fisher exact probability tests were carried out, but this time the number of questions in the partially correct and correct categories were added together. However, again the tests showed an identical pattern (all $p < 0.001$).

Between Task Transfer

The mean number of correct trials for each set of subjects in each group, for the entire transfer phase and each half of the phase can be seen in Figure 6.8. To explore how subjects coped when their task changed a 2 (task order) by 2 (learning goal) by 2 (phase score) analysis of variance with repeated measures on the last factor was used. This revealed a main effect of learning goal, $F(1,44) = 16.35$, $p < 0.001$: overall, performance in the test phase and the transfer phase was better for subjects with a pattern search goal. There was also a main effect of phase score, $F(1,44) = 5.85$, $p = 0.02$: subjects performed worse in the transfer phase than they did in the test phase. However, there was an interaction between phase score and learning goal, $F(1,44) = 17.67$, $p < 0.001$. Unlike in Experiment 6a, there were no other significant effects or interactions. To explore the interaction of phase score by learning goal, data were collapsed over task order and comparisons between test and transfer phase were made for the different learning goals. Paired sample t-tests showed that subjects with a control task goal had a significant drop in performance when they changed to their new task, $t(22) = 4.70$, $p < 0.001$. For subjects with a pattern search goal there was no drop in performance when the tasks changed. Hence, transfer of learning was only seen for subjects with a pattern search goal. This transfer occurred irrespective of task order.

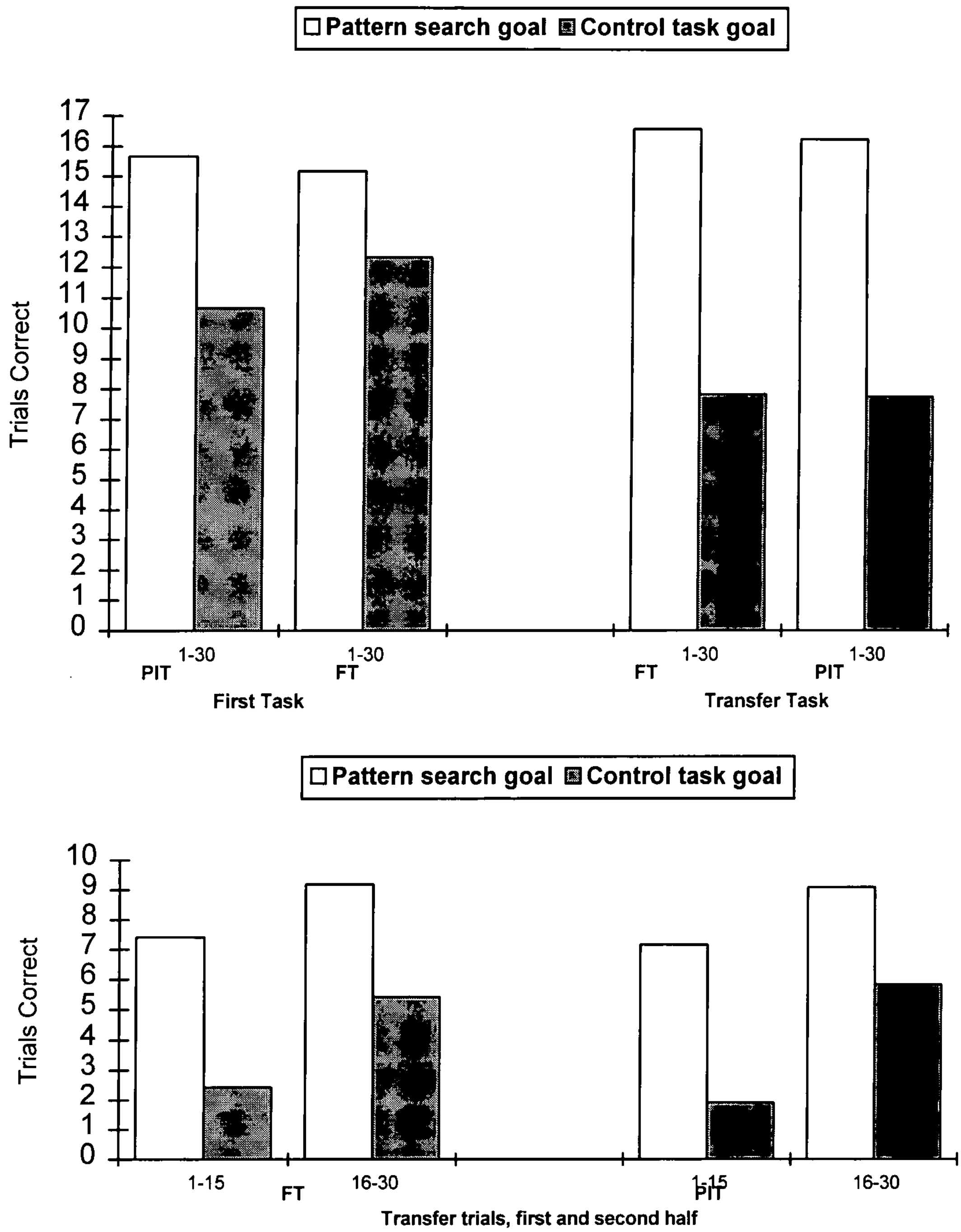


Figure 6.8: Mean number of correct trials in the transfer phase (and test phase for total scores) for each group. Data are shown for all 30 trials combined in the top graph and for the first and second 15 trials in the bottom. PIT = person interaction task, FT = factory task.

Correlation of transfer phase with earlier phases: For the pattern search group, Spearman Rank correlation coefficients were 0.72 ($p < 0.009$) for the subjects that switched from person interaction to simplified factory task and, 0.92 ($p < 0.001$) for the subjects that switched from factory to person interaction. For the control task subjects switching from person interaction to simplified factory task, the coefficient with the test phase was 0.21 ($p > 0.5$) and with the learning phase was -0.36 ($p > 0.2$). For the control task group switching from simplified factory to person interaction task, the coefficient with the test phase was 0.20 ($p > 0.5$) and with the learning phase was -0.10 ($p > 0.7$).

Opinions On The Similarity Of The Two Tasks

The average rating for each similarity type for each set of subjects within the two groups can be seen in Table 6.5 (the scale runs from 1 = extremely different up to 5 = extremely similar). The data in Table 6.5 were analysed using a 2 (learning goal) by 2 (task order) by 3 (type of similarity) analysis of variance with repeated measures on the last factor. The results revealed a significant main effect of learning goal, $F(1,44) = 27.89$, $p < 0.001$: on average for the three similarity types, subjects with a pattern search goal gave the two tasks a higher rating of similarity than subjects with a control task goal. There was also a main effect of type of similarity, $F(2,88) = 33.42$, $p < 0.001$. The interaction between type of similarity with learning goal just failed to reach significance $F(2,88) = 2.44$, $p = 0.093$. There were no other significant effects or interactions. With the interaction between type of similarity and learning goal nearly being significant, with data collapsed over task order, within groups comparisons were made between the different question types. For subjects with a control task goal the comparison between ratings for underlying similarity and strategic similarity just failed to reach significance, $Z = -1.68$, $p = 0.094$, suggesting a tendency for a higher rating of similarity for Strategic Similarity. Both Underlying Similarity and Strategic Similarity were given a higher rating of similarity compared to Overall Similarity. For Underlying Similarity vs Overall Similarity, $Z = -2.27$, $p = 0.02$ and for Strategic Similarity vs Overall Similarity, $Z = -2.83$, $p = 0.005$. Subjects with a pattern search goal gave a comparable rating of similarity for Underlying and Strategic and a higher rating of similarity for both these questions compared to Overall Similarity. For Underlying Similarity vs Overall Similarity, Z = -3.82, $p < 0.001$ and for Strategic Similarity vs Overall Similarity, Z = -3.62, $p < 0.001$.

Table 6.5
Similarity ratings of the two tasks for each set of subjects

PIT = person inter. task first FT = factory task first	**Category**	Overall Similarity	Underlying Similarity	Strategic Similarity
Task order	**Learning goal**			
PIT→FT	control task	**1.58**	**1.92**	**2.08**
PIT→FT	pattern search	**3.08**	**4.00**	**4.00**
FT→PIT	control task	**2.08**	**2.58**	**2.92**
FT→PIT	pattern search	**3.00**	**3.83**	**4.00**

Simplified Version Of The Factory Task vs The Old Version Of The Factory Task

This section of the results makes appropriate comparisons with Experiment 6a that the affect of the change in the factory task to be explored. The F tables of these analyses can be seen in Appendix 6, pg. 259.

Control performance during the learning and test phases of the experiment: Initially comparisons were made to answer the question of whether the simplified version of the factory task is easier to control than the normal version if it is a subject's first task. Since subjects with a pattern search goal had no performance measure during the learning phase, the different learning goal groups were explored separately. Starting with the subjects given a control task learning goal: A 2 (factory task: normal vs simplified) analysis of variance was used on scores for the total, first and last halves of the learning and test phases and on the New, Old-wrong, Old-correct type and total, prediction question scores. The results showed no significant effect of factory task on any of these scores. Apart from the learning phase scores, the same set of analyses were done for the subjects with a pattern search goal. Again, the results showed no significant effect of factory task on any of these scores. All in all, these results indicate that, during the initial parts of the experiment, ignoring learning goal, the simplified factory task was learnt and performed no better or worse than the normal version.

Control performance during the transfer phase: Transfer for the different versions of the factory task has been addressed in the results of the individual experiments. In the analysis of variances transfer of learning can be said to have occurred when there is *no* effect of phase score - subjects performance does *not* drop from the test phase of the first task to the transfer phase (the new task). The results have shown this to occur only for subjects with a pattern search goal. For subjects with a pattern search goal, transfer (lack of effect of phase score) has only been shown for the normal version of the factory task when subjects are switching *from* it. (The results from Experiment 6a showed an interaction between phase effect, learning goal and task order.) For the simplified version of the factory task transfer has been shown when subjects are switching both from it and to it. (The results from Experiment 6b did not show the three way interaction from Experiment 6a, but only showed an interaction between phase effect and learning goal.) The analysis reported here is to reinforce the finding of this pattern.

A 4 factor analysis of variance is used - 2 (learning goal) by 2 (task order) by 2 (factory task: simplified version vs normal version) by 2 (phase score: test vs transfer) - with repeated measures on the last factor. To use this 4 factor analysis of variance to reinforce these findings some complex interactions are expected and these need some explanation. Firstly, there should be a general effect of phase score indicating that subjects' performance is dropping overall between test and transfer phase. This indeed is the case; main effect of phase score, $F(1,88) = 15.52$, $p < 0.001$. Then there should be an interaction between this effect and learning goal as it is suggested that instance learners (those with a control task goal) should not show transfer (and hence should show an effect of phase score) and explicit learners (those with a pattern search goal) should show transfer (and hence *not* show an effect of phase score). This is also the case; for the interaction between phase score and learning goal, $F(1,88) = 15.92$, $p < 0.001$. This last interaction was also significant in the second experiment (which had the simplified factory task) and not in the first experiment. Therefore in the 4 factor anova reported here, the crucial expectation is a three way interaction between phase score, learning goal and factory task. This interaction nearly reaches significance, $F(1,88) = 2.83$, $p = 0.096$.

Additional to these interactions the three way interaction between phase score, task order and factory task is significant, $F(1,88) = 4.13$, $p = 0.045$. This represents the tendency for the effect of phase score and task order to interact when the factory task is the normal version, but not when it its the simplified version. The three way interaction between phase score, learning goal and task order is also significant, $F(1,88) = 5.22$, $p = 0.025$. This represents the tendency in Experiment 6a for the interaction of phase score by learning goal being dependent on which task was done first.

Similarity ratings of the two tasks: Next comparisons were made to answer the question of how the simplified version of the task effects the recognition of the underlying similarity between person interaction and factory tasks. In order to do this a 2 (learning goal) by 2 (task order) by 2 (factory task) by 3 (type of similarity) analysis of variance with repeated measures on the last factor was used. This revealed a main effect of learning goal, $F(1,88)$ - 38.31, $p < 0.001$: subjects with a pattern search goal were more likely to say the tasks were similar than subjects with a control task goal. There was also a main effect of type of similarity, $F(2,176) = 47.01$, $p < 0.001$ and a significant interaction between type of similarity and learning goal, $F(2,176) = 5.93$, $p = 0.003$. There were no other significant main effects or interactions. The lack of any significant results involving the effect of factory task suggest that whether the factory task was normal or the simplified version did not affect how subjects rated the similarity of the two tasks.

DISCUSSION

Firstly the main conclusions one can make from the results of Experiment 6b are summarised. With regards to mode of learning, again, type of learning appeared to be determined by the type of goal set during the learning phase. Subjects with a pattern search goal learnt through explicit hypothesis testing irrespective of whether the task was the person interaction task or the simplified factory task. Subjects with a control task goal learnt through instance learning irrespective of which the task was. With regards to transfer of learning, as predicted, for subjects with a pattern search goal, switching from the person interaction task to factory task, learning did transfer. This key result did not occur for Experiment 6a's equivalent pattern search subjects. Similar to Experiment 6a, learning did transfer from the simplified version of the factory task to the person interaction task for the pattern search subjects. Also similar to Experiment 6a learning did not transfer in either direction for the control task subjects. A more detailed consideration of the data explains how these deductions have been made:

Type of learning : As in Experiment 6a, it is deduced that the pattern search subjects learnt explicit rules from the following points: (1) The similarity of performance on all prediction question types, particularly, being able to do just as well at new questions as at the Old-wrong and Old-correct questions; (2) The success at answering the rule description questions; (3) The significant positive correlations between total questionnaire scores and test phase trial performance. Evidence for the control task group learning through instance learning is inferred from the following points: (1) The significant difference in performance on the different prediction question types, namely, poor performance on the new questions; (2) The poor answers to the rule description questions; (3) The lack of a significant positive correlation between total questionnaire score and trial performance. Again, the notion of the two groups learning differently is further reinforced by the comparisons between the groups; the pattern search group performed better on both the prediction questions and the rule description questions, indicating that they had acquired significantly more verbalisable knowledge than the control task group. These results from Experiment 6b add further support to the hypothesis that groups with a pattern search goal should learn explicit rules while groups with a control task goal should learn through instance learning. The results now clearly indicate that, in addition to the other tasks, this hypothesis holds true for the previously untested simplified factory task.

Transfer of learning: for the groups given a pattern search goal learning transferred from the first task to the second irrespective of the order of the tasks. This is deduced from the following points: (1) The comparable performances between test and transfer phase on the overall scores. This indicated that there was no overall drop in control ability between the two tasks, hence what had been learnt in the first task carried over to the second. (2) The significant positive correlation between test phase and transfer phase performances. This suggests a close link between performance in the test phase and performance in the learning phase. As suggested in Experiment 6a, it is likely that control methods used by subjects in the test phase were then transferred to the transfer phase. For the control task group it is deduced that learning on the first task did not successfully transfer in any way to the second task (what ever it might be) from the following two points. (1) The significant lower levels of performance during the transfer phase when compared to the test phase. (2) The lack of significant correlation between transfer phase performance and performance in either the learning or the test phases. This indicates that performance in the first task was not related to performance in the other. In summary, the transfer of learning results describe a pattern of (a) transfer of learning occurring for the pattern search group, b) no transfer of learning occurring for the control task group.

Recognition of the underlying similarity of the two tasks : As for Experiment 6a, it appears that generally those subjects learning instances were less likely to label the two tasks as similar than those subjects learning explicit rules. This can be deduced from the fact that there was an overall effect of learning goal on type of similarity: subjects with a pattern search goal outperformed those with a control task goal. However, within group comparisons revealed a similar pattern for the control task goal subjects of Experiment 6b to that of the pattern search subjects from both Experiment 6a and 6b. The control task subjects also appeared to have their greatest recognition of the similarity of the two tasks coming from recognising the underlying and strategic similarities. Perhaps the simplified version of the factory task allows this to be more transparent. Still though, the pattern search subjects gave significantly higher ratings of similarity than the control task subjects.

Comparisons of the simplified version of the factory task with the original version show no difference in performance during the pre-transfer stage of the experiment. The lack of difference in the pre-transfer phase is important. It means that any evidence of transfer to the new task is not due to subjects having reached a higher competency of control in the initial task, but is due to some more intricate difference between the normal and simplified versions of the factory task.

GENERAL DISCUSSION

Aside from the transfer of learning, the two experiments presented in this chapter have replicated and extended the work of Experiment 1 that illustrated the relationship between the goal of a learning task and subsequent mode of learning. Experiment 1 emphasised that on the person interaction task an initial non control task goal leads to explicit hypothesis testing and rule deduction whereas a control task goal leads to instance learning. In this study, the range of this finding has been extended. It has been replicated for the person interaction task and has also been demonstrated to hold true for the original Berry & Broadbent's sugar production task and on a modified version of that task (all using the same underlying equation).

Another element of Experiment 1 that has been replicated is the rule learners of the person interaction task significantly outperforming the instance learners when controlling the task. This result suggests that not only does explicit rule learning lead to better declarative knowledge, but also leads to better control performance. This result was replicated for both the normal and simplified versions of the factory task.

Far transfer of learning was not seen for the control task instance learners. It must be remembered that the tasks to which subjects were switching had an identical underlying structure to the task they were switching from. Any model of instance learning must take into account the fact that, despite an identical underlying abstract pattern between the two tasks, success on the first task did not transfer to the second. As expected, this indicates that instance learning mechanisms are dealing with the surface elements of the task. Change the surface elements and what has been acquired through the instance learning mechanism is of no use in the subsequent task. Importantly, the lack of transfer of learning therefore suggests that any instance learning mechanism is not abstracting some then universally applicable underlying pattern. A look-up table model of instance learning would account for this lack of transfer. The look-up table consists of instances or 'exemplars' of the task. Future problems based on the task relate to the look-up table model to see if they have encountered them before. If a new task was encountered that only had an identical pattern to the old, but different surface structure, the old look-up table would be no use in coping with new task as it has no abstracted information concerning the underlying pattern of the two tasks. Obviously, in this present study, there is evidence for the look-up table's feature of a low level of transference ability.

Even when the surface structure remained the same and simply the specific in-task goal changed, control task subjects' performance drops. Evidence for this can be seen when the subjects change specific goal between learning and test phases. However, also as expected from learning that could be modelled by a look-up table, some transfer does occur as overall, subjects' performance is comparable between learning and test phases. This is a demonstration of near transfer. The measures needed to explore near transfer were also made for all the other experiments, (see pages 28, 55, 86, 115, & 137 for the transfer results of each experiment). Considering transfer is the focus of study of this chapter, it is appropriate to consider those results here and how they reflect on the findings of this chapter.

To summarise, all the results from the other experiments support the notion that instance learners can demonstrate near transfer. If anything, the results of the other experiments show more evidence of near transfer than is shown from Experiments 6a and 6b. Comparisons with the results of Experiments 3 and 4 are confounded somewhat by having secondary tasks occurring while trying to achieve the specific goals, however the results for these two experiments are in the same direction as that of experiments 1,2 and 5 that are properly comparable to the two experiments in this chapter. The only other difference with the two experiments in this chapter is that, excluding Experiment 2, the other groups' comparisons of performance across learning and test phases have the additional dual goal subjects results in the design. These are instance learners too so the examination of near transfer is also appropriate for them. However, the ANOVAs making the comparisons have learning goal as a factor so the examination of near transfer can be made just for the control task subjects as in the two experiments in this chapter. As regards how subjects coped just after the task changed (explored by the comparison of performance during last half of learning phase vs performance during first half of test phase), unlike the two experiments in this chapter, none of the other experiments showed a drop in performance. This provides more evidence of near transfer. As regards how subjects coped overall between learning and test phases, Experiment 1 shows the same as that of the two experiments in this chapter (subjects performed comparably between the two phases). Experiments 2 & 4 showed a significant increase in performance for the test phase and experiments 3 and 5 showed tendencies for an increase in performance . All in all the results of the whole thesis, as expected from a look-up table model of instance learning, show evidence for instance learners showing near transfer.

This result of instance learners displaying no far transfer (when the surface structure changes), but near transfer (when the surface structure remains the same and only the specific goal has changed) is consistent with the studies of near transfer discussed in the introduction of this chapter. Both Anderson's (1983) ACT model and Logan's (1988) instance model assume that learning involves the memorisation of productions and instances respectively. Hence, matching the data of this study, transfer only occurs when the learning and transfer tasks are perceptually similar, as well as structurally similar.

Transfer of learning in appropriate conditions should be expected for explicit rule learners. As illustrated in the introduction of this chapter, the pattern search goal given to some of the subjects should have some chance of leading to far transfer. The evidence of this study adds to the body of evidence detailed in the introduction of this chapter (e.g. Gick and Holyoak, 1983; Owen and Sweller, 1985; Sweller, Mawer and Ward, 1983; Vollmeyer, Burns and Holyoak, 1996) that demonstrates far transfer. Learning transferring for the pattern search subjects (under certain conditions) is further evidence that a non control task, pattern search goal leads to explicit learning. The fact that learning never transferred for the control task subjects helps emphasise how differently these subjects learnt from those with a pattern search goal. It reinforces the idea that specificity of goal has a determining effect on the way people learn.

Transfer of learning for the explicit rule learning subjects needs some scrutiny. It only occurred in some conditions. It was seen to occur in either direction for the simplified version of the factory task, but only *from* the normal version of the factory task to the person interaction task. Models of explicit rule learning that allow for transfer of learning need to explain this. The problem appears to be solely related to the mechanics of transfer as the results suggest that for the pattern search subjects, the person interaction task is no easier or harder than either version of the factory task. A few key points must be born in mind. The simplified version of the factory task was designed so that mapping from the person interaction task to the factory task would be easier as both tasks would have an identical number of variables and 'currencies' to consider. This obviously aided transfer. The fact that transfer would only work *from* the normal version of the task is another important point to examine. The normal version of the task has more variables for the learners to consider than the person interaction task. It appears that subjects have to understand this more complex task first then a switch to a task with fewer variables to consider is competently done, but not the other way round. These findings are similar to that of other studies. Hayes and Simon (1974), using their 'monster problem', showed that subjects had difficulty transferring to a second problem when the second problem, although almost identical to the first, had subtle structural differences related to one of the variables. Reed, Ernst and Banerji (1974), using their missionaries' wives problem, showed that subjects only showed transfer between two almost identical problems, when subjects initial learning was on the harder of the two (even then subjects needed a hint to identify that the two problems were similar). Bassok and Holyoak (1993), showed that training on an equation, set in an algebra class context, transferred to use of the same equation in a physics class context, but not vice versa. As in the study presented in this chapter, Bassok and Holyoak suggested that the reason the transfer was asymmetrical was due to less than optimal mapping of elements between tasks caused by "....a discrepancy between the quantity types of the key variables in the source and target problems (pg. 93)".

One other possible reason why transfer from the normal version of the factory task to the person interaction task works but not the other way round is a key difference between the two tasks. The factory task, unlike the person interaction task deals in numerical inputs and outputs. Possibly encountering this sort of task first allows a more abstract underlying pattern to be established and this makes transfer easier.

In summary, this paper has explored how learning transfers between two tasks that have an identical underlying structure but a different surface one. The paper has explored this potential transfer of learning for both groups of subjects who are learning as instance learners and those who are learning through explicit hypothesis testing and rule deduction. Instance and rule learning of the same tasks was induced by having either a specific control goal or a non-specific pattern search goal. First it was shown that the different types of learning had been induced by this method. Then transfer of learning was explored. It was shown that transfer of learning did not occur for those groups of subjects learning instances. This suggests that instance learning mechanisms do not abstract an underlying pattern that they can transfer from one task to another. Transfer of learning did occur for the explicit rule learners, but only when the task to which subjects transferred had fewer or an equal number of variables to consider.

Discussion

The final discussion of this thesis is organised as follows: Firstly, there will be a brief summary of the main conclusions. Following this, the theoretical issues and implications raised by the thesis will be considered. The results of the three learning goals and the related learning modes will be detailed and examined in turn and their theoretical implications considered. Following this the effectiveness of the dual space model to explain the learning goal effect will be discussed. Also, the models used to describe implicit learning will be recalled. Finally, Reber (1993) has also suggested some particular differences of an evolutionary nature that should occur between modes of learning that should be directly testable in the studies presented in the thesis. Therefore, there will be some discussion about his ideas and how the results reflect upon them. Then, methodological issues and implications will be discussed. The prediction questions will be the main focus of this section. A potential flaw in their design, raised by Experiment 5 (the memory experiment), will be examined because this could have confounding effects on the results. However, it will be demonstrated that this potential flaw does not actually affect the results. Additionally the exact nature of the prediction questions as a learning mode indicator will be scrutinised. Then, the practical issues and implications of the results will be discussed. Answers will be provided to the following questions; What practical implications do the results have for training courses?; What do the results tell us about the potential advantages of varying goal specificity?; Has the learning studied in the thesis been too lab based to allow the results to have useful practical implications? Finally the discussion will be summarised with particular emphasis on listing recommended future work that should be carried out.

Summary Of The Main Conclusions

Experiment 1, the learning goal effect experiment, suggested that a pattern search goal leads to explicit rule learning, a control task goal leads to instance learning where subjects memorise both correct and incorrect trials, while a dual goal, (a combination of control task and pattern search goals), leads to instance learning where subjects memorise correct trials only.

Experiment 2, the observation experiment, demonstrated that the learning goal effect did not arise because the goals induced subjects to view goal specific ranges of information (the 'salience explanation'). Whilst this was a feasible explanation, the results showed that pattern search subjects who observed control task models still learnt rules in the same way as had previous pattern search subjects. Therefore the 'goal explanation', that is, that learning goals affect learning by directly influencing cognitive processes, was confirmed as the likely explanation of the learning goal effect.

Experiment 3, the concurrent verbalisation experiment, explored the 'goal explanation' by examining what direct cognitive effect learning goals may be having. It was concluded that a pattern search goal leads to hypothesis testing and the exploration of rule space, whereas a control goal leads to instance learning, because it promotes and confines subjects to the exploration of instance space only. The experiment also demonstrated the positive effects of explanations on learning for both rule and instance learners.

Experiment 4, the random number generation experiment, demonstrated that the combination of control task and pattern search goals leads to a cognitive load that is too high for explicit learning processes to take place. Therefore the pattern of data of the dual goal subjects reflects pure implicit learning. When subjects had an additional, task that occupied the central-executive of working memory (random number generation), the pattern of performance for all goal groups was the same as that of the dual goal subjects from the learning goal effect experiment.

Experiment 5, the memory experiment, revealed that pattern search subjects may be learning instances as well as rules. Additionally, it appears that though dual goal subjects' post learning task performances indicate that they do not use incorrect instances, they still have some memory of these instances. The control task subjects' memory performance matched that expected from their post learning task performances.

Experiments 6a and 6b, the transfer experiments, demonstrated that explicit instance learners, that is control task subjects, could perform near transfer, i.e. transfer of learning between two identical tasks with different specific goals. However, they could not transfer their learning to a task that, though structurally identical, was perceptually different. On the other hand rule learners, that is pattern search subjects, could transfer to a novel task but only if the novel task was of equal or of a lower complexity than the task subjects transferred from.

Theoretical Issues And Implications

The control task goal: implicit/explicit instance learning

Issues and implications relating to learning caused by a control task goal are now examined. The results for subjects with a control task learning goal showed that they learned instances. Some of the evidence for this comes from the rule description questions as subjects could not describe the underlying rule that Clegg followed or explain what principles they used to make Clegg behave the way they wanted him to. The other evidence comes from the fact that the subjects could not make predictions from novel situations, indicating that they had no abstract rule of the task. Their prediction question performances showed that they were equally good at making predictions from correct familiar situations and from incorrect familiar situations. These results were replicated in all of the experiments (excluding the random number generation experiment where, due to the random number generating condition, different results were expected).

The results of the observation experiment and the concurrent verbalisation experiment taken together indicated that the control task goal leads to instance learning due to a direct influence on cognitive processes, which comes in the form of confining subjects to the search of instance space. This is an important finding as previously it was thought that it was the low salience of the material that caused instance learning. In other words it was thought that instance learning only occurred because the material was too complex. The implication behind such an assumption is that subjects were attempting to find a pattern but could not, because it was too difficult to deduce. The learning goal effect refutes this assumption. Clearly, the task is not so complex as to prevent the learning of its underlying rule because pattern search subjects can learn it. It is the requirement of controlling the system to achieve a specific goal that leads to instance learning because the requirement confines subjects to a search of instance space. Therefore, contrary to the implicit assumption, subjects have either not even been attempting to determine the pattern or if they have, they have not been able to make any headway as their goal confines them to the search of instance space only.

However the learning goal effect does not rule out salience as a relevant factor in determining whether people learn rules or instances. It is just that in the Clegg version of the person interaction task, the learning goal is a more fundamental determining factor. As reviewed in the learning goal effect experiment, there is plenty of evidence to demonstrate that salience affects learning (e.g. Berry & Broadbent, 1988; Hayes & Broadbent, 1988). One important point that further work may want to examine is how the learning goal effect and the factor of salience interact. Any practical implications of the learning goal effect, regarding the structure of training programmes or regarding teaching / problem setting strategies, may depend on the salience of the task. Given processing constraints due to the cognitive capacity needed for hypotheses creating and testing, it is quite likely that when a task gets too complicated, pattern search instructions may not lead to explicit rule learning. Exactly when and if this happens and what are the best ways to set up training problems relative to the learning goal and the salience of the material to be learnt, needs to be established. To examine the interaction of the learning goal effect with the factor of salience a precise method of altering and indeed measuring salience would be needed. The present method used by experimenters (e.g. Berry & Broadbent, 1988) does not allow an exact measure of the level of salience of material to be made. Ideally a system would be devised so that a task could be classified along a scale. With this, any differential effect of the learning goal effect would be easier to pin down.

The results of the random number generation experiment demonstrated that part of the control task subjects' learning occurs explicitly. When control task subjects had a secondary task of random number generation they could no longer make predictions in Old-wrong situations, only in Old-correct situations. The idea that control task subjects are performing instance learning through a mix of implicit and explicit processes is in keeping with the findings of Dienes and Fahey, (1995) who adopted Logan's instance theory to explain their results. As detailed in the introduction of Chapter 6 (see page 156), Logan's instance theory assumes that subjects start off using explicit strategies, but at the same time, they accumulate instances in memory.

Once subjects have accumulated enough instances these are then recalled quicker than the explicit strategies and so instance learning is dominant. Hence Logan's theory, in keeping with the results of the control task subjects of this thesis, suggests that instance learners are using a mixture of explicit processes (the initial use of strategies) and implicit processes (the memorisation of instances).

Initially it was proposed that the learning of the Clegg version of the person interaction task occurred through purely implicit processes. One of the main justifications for this was because subjects performed at chance on prediction questions (e.g. Berry & Broadbent, 1984). However, the results of this thesis suggest that chance performance on prediction questions in previous studies was only because of the way their prediction questions were designed. The previous studies made no attempt to categorise prediction question types. The prediction questions need to be designed so that they have a discrete set of New, Old-wrong and Old-correct questions. It is only with these carefully designed prediction questions, (and then using methodologies like in Experiment 1 and Experiment 4), that the conclusion that control task subjects' instance learning is actually a combination of implicit and explicit processes can be teased out. (Previous research has also assumed that poor performance on prediction questions means subjects have no explicit knowledge of the task. That is, they have assumed that prediction questions only tap explicit knowledge - this particular point is addressed below in the section on Methodological issues.)

The conclusion that instance learning caused by a control task goal does indeed have some element of explicit processes confirms Shanks and St. John's (1994) notions that there was insufficient evidence from previous studies to declare that subjects were learning implicitly. They concluded that it is best to simply refer to the learning as instance learning and assume it is explicit until sufficient evidence is supplied to the contrary. One of their main claims was that the measures used were not sensitive enough to detect whether subjects had any explicit knowledge. The results of the thesis confirm that claim, since the studies presented here have demonstrated that only with discrete sets of Old-wrong and Old-correct prediction questions can it be concluded that subjects do have an explicit component in their learning.

The memory experiment revealed that control task subjects do have both Old-wrong and Old-correct instances stored in memory. The experiment also showed that subjects were particularly good at detecting instances that did not follow the underlying rule that Clegg followed (that is the New-illegal instances). One possibility put forward was that subjects were using a simple guiding principle to make these detections. This could be something such as simply deciding that any instance that contained an extreme emotion such as *Loving* or *Very Rude* was not one they had encountered. Importantly, this idea would not call for subjects to encode their instances in any abstract manner as use of such a principle relies only on the surface features of the instances and does not require any extra information to be retrieved from memory. The results of control task subjects from the transfer experiments can be taken to add tentative support to the notion that a simple guiding principle was used to detect New-illegal instances and therefore instances do not need to be encoded in an abstract manner. The transfer experiments showed that control task subjects were not able to demonstrate far transfer. To display far transfer, subjects would need some memory for abstract information. Since they do not display such a memory, it is less likely that they encode instances in any abstract fashion and therefore it is more likely that a simple guiding principle, such as detection of an extreme response, was used to detect New-illegal instances.

The dual goal: implicit instance learning

The subjects with a dual goal clearly learned instances. Like the control task subjects they were poor at answering the rule description questions and could not make predictions from novel situations. Further, to allow them to fall into the category of instance learners, they did demonstrate the ability to make predictions from familiar situations. However, in contrast to control task instance learners, they were able to make predictions from correct familiar situations better than from incorrect familiar situations. This pattern of results was replicated in all of the experiments (excluding the transfer and observe experiments where dual goal task conditions were not included).

The random number generation experiment in combination with the learning goal effect experiment supply strong evidence for the dual goal subjects having learned purely implicitly. In the random number generation experiment a secondary task (random number generation) was designed to occupy the central executive of working memory - a core ingredient for explicit learning processes. The evidence suggested that occupation of this mechanism was successful, therefore, any abilities that remained should be gained through implicit learning processes. All subjects in the random number experiment were still able to perform well on Old-correct prediction questions. Notably, all subjects, irrespective of their learning goal, performed comparably to the dual goal subjects from the learning goal effect experiment. It was concluded that, as hypothesised, dual goal subjects without a secondary task do not learn through explicit processes either and therefore they are learning purely implicitly. In other words, correct trials are learned implicitly.

Shanks and St. John suggest that nearly all implicit learning experiments attempting to demonstrate implicit learning have adopted a *single* dissociation logic by attempting to demonstrate learning in the absence of any awareness. They point out that the problem with this procedure is methodological in that the tests that demonstrate the absence of awareness are arguably not sensitive enough to exhaustively demonstrate that no explicit awareness exists. As discussed above and at length in the random number generation experiment, they make a valid point. In the case of dual goal subjects, the old style prediction questionnaire frequently used to support the claim of implicit learning, would not have been sensitive enough to demonstrate that dual goal subjects could make predictions from certain situations. Therefore the old style prediction questionnaire would not have been sensitive enough to tease out the exact nature of dual goal subjects' learning.

The results from this thesis have a number of advantages over the commonly used procedure for demonstrating implicit learning that simply attempts to demonstrate learning in the absence of any awareness. For one, the tests used in the thesis are more sensitive than the ones used in other studies. Apart from the already mentioned improvements made to the prediction questions, the marking of the rule description questions was done in both a liberal *and* *a conservative* manner to make absolutely sure that goal conditions were not demonstrating explicit knowledge. The conservative condition allowed answers to be counted as correct even if they were only partially correct. This conservatism means that subjects with partial knowledge of the rule were not penalised by the scoring procedure, since only completely wrong answers were taken to indicate no explicit knowledge. Additionally the process used in the random number generation experiment concentrated on completely occupying the central executive of working memory, thus, relative to other studies, making it more certain that explicit processes were not occurring.

The use of the random number generation task means that there is one particularly important point that distinguishes the experiments in this study with other dynamic systems studies that demonstrate implicit learning. It means that the work of this thesis goes beyond simply demonstrating learning without awareness. The studies presented here demonstrate qualitative differences between learning modes. The random number generation experiment shows that the implicit instance learning mode is not sensitive to a central executive occupying task whereas the control task instance learning mode and the explicit rule learning mode are. Demonstration of qualitative differences makes the argument stronger and as suggested by Shanks and St. John (1994, pg. 369), leads to the evidence for implicit learning being firmer.

The demonstration of qualitative differences between learning modes has been attempted before by trying to demonstrate that learning modes are differentially sensitive to salience (e.g. Berry and Broadbent, 1988; Hayes and Broadbent, 1988). The problem with these studies is that difference in performances between subjects supposedly adopting different modes of learning can always be attributed to differences in the complexity, or salience, of the task that subjects are learning. In other words, subjects may be learning better because the task they are learning is less complex and not because they are learning explicitly. (For more detail of this argument see the introduction of the random number generation experiment, pg. 108.) The conclusions reached about the qualitative differences between learning modes in this thesis can not be explained away by such an argument as all subjects, whatever learning mode they were using, learned exactly the same task.

Shanks and St. John suggest that exceptional evidence for implicit learning requires a *double* dissociation where a particular pattern of qualitative differences between learning modes is demonstrated. For instance, mode X could be affected by one variable, whereas mode Y is completely unaffected, and then mode Y could be affected by another variable whereas mode X is this time left unaffected. There is probably a method to do this with the techniques used in the observe and random number generation experiments. It has already been shown that a central executive occupying task such as random number generation will affect rule learners and explicit instance learners, but not implicit instance learners. To meet Shanks and St. John's criteria for exceptional evidence, a condition needs to be established that will impact on implicit instance learning, but not explicit rule learning. The observation paradigm used in the observe experiment may be ideal for this. Observing during the learning phase did not affect explicit rule learners - they performed as expected. However, it did affect instance learners who learned nothing (they performed at chance on all measures). The instance learners were only given control task instructions not the dual goal instructions, so to get the double dissociation condition between implicit instance learners and explicit rule learners the dual goal condition would need to be run in a new experiment. If their results turned out to be the same as the control task condition (i.e. no learning), then the results, taken in conjunction with the learning goal effect experiment and the random number generation experiment, would meet the double dissociation condition for exceptional evidence of implicit learning.

There is good reason to speculate that the dual goal observers would learn nothing. The observation paradigm removes the possibility of action during the learning phase. It was suggested in the random number generation experiment that the direct link with an instance and action that occurs for correct responses may be the reason why implicit learners can make predictions from Old-correct situations. If the possibility of action was removed then it is predicted that no implicit learning would transpire as there would be no situations for direct links with action to occur. If an experiment showed that dual goal observers could not learn, then it would be possible to conclude that for implicit instance learning action is vital.

There is one alternative possibility as to why implicit instance learning may not occur in such an experiment. It may be that it is actually feedback that is vital to implicit learning, not action. It may be the lack of feedback that was preventing control task observers from learning. The instructions provided to the observers gave no information about the goal of the model subjects and so there was no information about whether the model's responses were correct or incorrect. It is quite possible that with this extra information the results of the control task would have been different. Any experiment looking for the double dissociation, attempting to show that action is vital to implicit instance learning, but not to explicit rule learning, would also have to examine the importance of feedback as this may be what would prevent learning on the part of implicit instance learners. If feedback, not action, turned out to be the vital variable for implicit instance learning, then the goal of a double dissociation may not be attainable. It is hard to see how feedback would not also be needed for explicit rule learning. However, it may be a certain sort of feedback that is vital for implicit instance learning, but not for explicit rule learning, in which case the double dissociation would be attainable. Exactly how to define the various sorts of feedback will need to be examined by further work. However, it seems likely that the feedback needed for rule learning is information about the relationships between input and output, while the feedback needed for instance learning is information about whether or not a given response is correct.

The memory experiment showed that dual goal subjects, though having a better memory for Old-correct instances, still had a good memory of Old-wrong instances. As was discussed in the memory experiment, this causes problems for the assumption that a look-up table, used to explain instance learners' post learning task abilities, represents the sum total of a person's memory. This assumption stems from the notion that the look-up table is simply a reflection of memory. In the discussion of the memory experiment, one of the proposals used to explain this finding was that implicit instance learners may have encoded memories more strongly for Old-correct instances and therefore only these stronger memories are entered into the look-up table. There is a way to examine this proposition. The experiment could be re-run, however the recognition test could be modified so that the recognition of each instance could be accompanied by a confidence rating question. If there is some truth in the above proposal, then it would be expected that dual goal subjects' confidence ratings for recognition of the Old-correct instances should be higher than ratings for the Old-wrong instances. If it were established that only strong memories are entered into a look-up table, it would be clearer how to refine a look-up table model so that it can explain both implicit instance learners' post learning task abilities and the information they have stored in memory.

The pattern search goal: explicit rule learning

The subjects with a pattern search goal explicitly learned the rules that Clegg followed. Apart from the fact that the subjects could correctly state the rules, they were also able to make predictions from novel situations as well as from familiar situations. Additionally, there was a significant positive correlation between prediction performance and control performance - what would be expected if subjects were using their knowledge of the rules to both make predictions and control the system. These patterns of results were replicated in all of the experiments (excluding the random number generation experiment).

There were other results from the pattern search subjects that would be expected exclusively of explicit rule learners. In the observation experiment, the lack of action caused by the observation condition, debatably vital to subjects simply performing instance learning, did not interfere with the pattern search subjects' pattern of learning. In the concurrent verbalisation experiment, the describing condition, compatible with instance learning, dramatically interfered with the pattern search subjects' normal pattern of learning and clearly prevented rule learning. In the transfer experiments, far transfer was only shown for pattern search learners, an ability that is arguably definitive of rule learning.

The only unusual result for the pattern search learners was from the memory experiment that showed that, contrary to the findings from other studies (e.g. Barclay, 1973; Nosofsky et al, 1989), the pattern search learners appear to have memories of instances. It is questionable whether they also had memories for instances from their learning phase where they only had a pattern search goal or just from their test phase where they had a specific control goal. The reason it is questionable is that subjects were only tested on whether they had a memory of test phase instances. An obvious way to redesign the experiment so that it could be seen whether pattern search subjects also learnt instances during their learning phase is to have the recognition test straight after the learning phase making the Old instances in the recognition test come from the learning phase. This way the test would be examining subjects' memories of instances when they have only had a pattern search goal. It would therefore prove conclusively whether or not a pattern search goal on its own also leads to instance learning.

It is important to establish whether pattern search subjects do learn instances as it would then need to be established how important their memories of instances are to their learning of rules. For example, one possibility is that these subjects do learn instances and only later induce a rule on the basis of the learned instances. The results of the memory test suggest that memories of instances are important in at least one way. These subjects appear to use their memories of instances to make recognition judgements, not rules they have learned. This was obvious from the fact that they did not incorrectly recognise New-legal instances, something that would have happened if they had been using their knowledge of the system's rules to make their recognition judgements. However, as mentioned in the memory experiment, the apparent strength of instance use over rule use to make recognition judgements may not have occurred if the instances subjects were recognising had more information, making the application of the rule easier.

There is already some evidence from the thesis that suggests that learning of instances is not vital to rule deduction. In the observe experiment the pattern search observers, observing a model subject's data who had been given a control task goal, still managed to deduce the rules of the system. The data that these observer subjects were viewing had a significantly narrower diversity of instances compared to normal pattern search subjects. If pattern search subjects were relying on instance learning to deduce their rules then it would be expected that being exposed to a narrower range of interactions would in some way impede their rule learning. As it did not, it is fair to assume that instance learning does not play a vital role in their explicit rule learning.

Further evidence that instance learning is unimportant to pattern search learners comes from the concurrent verbalisation experiment. The implication behind the notion that explicit rule learners are primarily learning instances, then inducing their rules from these instances, is that the importance of hypothesis testing is not as strong as it would otherwise be if instance learning was not important to rule deduction. The importance of hypothesis testing was clearly shown in the concurrent verbalisation experiment. When pattern search learners were prevented from hypothesising in the describe condition, then rule learning was also prevented. Also, when hypothesis testing was encouraged by the explain condition, then rule learning was considerably improved with all subjects being able to induce the rules of the system. All in all, it is likely that the evidence from the memory experiment suggesting that pattern search subjects have some memory trace of instances does not mean that instance learning is the fundamental learning process underlying rule learning.

Another study that could be carried out that would help clear up the problems of exactly how pattern search learners are learning and additionally provide information about how the other goals influence learning, would be one where undirected 'talk aloud' verbal protocols were collected. At present, the precise learning strategies that the subjects used for either instance learning or rule learning can not be absolutely determined. It would therefore be useful to collect verbal protocols while subjects are working towards different goals to determine the kinds of explicit strategies that they are using. One would expect the protocols to contain evidence of hypothesis testing strategies when subjects are given a pattern search goal, consistent with the findings of Vollmeyer et al (1996). With the control task subjects, one would expect the protocols to contain evidence of means-ends analysis or of other strategies subjects use when explicitly trying to reach a control task goal, consistent with the view that instance learning in the standard person-computer interaction task includes a mixture of implicit and explicit processes. Of particular interest would be the protocols of the dual goal subjects. As suggested by the random number generation experiment, these subjects are unable to learn explicitly, therefore, there should be little evidence of any explicit strategies, either for reaching the specific control goal or for testing hypotheses to reach the pattern search goal. However, Whittlesea and

Dorken showed that subjects can memorise instances as a consequence of the purpose of the task they are set. Therefore, with regard to the instance learning that occurs in the dual goal group, it might be discovered what it is the subjects are doing that causes them to memorise correct instances without apparent effort.

The dual space model

Klahr and Dunbar's (1988) dual space model suggests that subjects' learning can be thought to occur in one or both of two spaces - instance space and rule space. Subjects exploring rule space form hypotheses which they can then test in instance space. Klahr and Dunbar designed their dual space model to explain how scientists reason. It has been adopted in this thesis to explain how the learning goal effect leads to either rule learning or instance learning. Essentially, the model describes deliberate, or explicit, learning (either rule or instance), and the model is used in the thesis to explain how people's reasoning processes are *directly* guided by the different goals. In the case of subjects with a dual goal, the resulting implicit learning is an *indirect* effect of the cognitive load being too heavy. Thus the dual space model, which is crucial for explaining the learning goal effect, does not so easily explain the dual goal results. One possibility is that implicit instance learning also occurs in instance space, but this still leaves open the question of how implicit instance learning occurs. An alternative possibility is that implicit instance learning is independent of the dual space model. This possibility seems the most plausible, since the dual space model was only intended to explain explicit processes. The issue of how implicit learning might occur will be returned to after the dual space model has been discussed.

The other effect the dual space model was able to describe occurred in the concurrent verbalisation experiment. Exactly why should describing aloud be incompatible with a pattern search goal and compatible with a control task goal? The model dictated that the 'describe' verbalisation condition should restrict subjects to instance space which was essentially what a control task goal was doing therefore the conditions would be compatible. However, the 'describe' verbalisation condition would prevent the pattern search subjects from exploring rule space, so therefore it would be incompatible with the pattern search goal. The model was also shown to have advantages over models stemming from the concept learning literature that combine theory learning and empirical learning (e.g. Wisniewski and Medin's (1995) interactive model, see the concurrent verbalisation experiment pg. 99.)

The implicit-explicit learning literature has shown evidence for two distinct learning processes - instance and rule learning. Turning to another domain, the concept learning literature has also shown evidence for two distinct cognitive processes which are directly comparable to instance and rule learning. Indeed in the concept learning literature there are two distinct class of theories to explain cognitive processes (categorisation) based on rules and cognitive processes based on instances. For example, as classified by Komatsu (1992), the exemplar view (e.g. Medin and Schaffer, 1978) gives psychological accounts of how subjects use instances to categorise objects, and, the explanation based view (e.g. Johnson-Laird, 1983; Murphy and Medin, 1985's theory driven model) gives psychological accounts of how subjects use rules to categorise objects. Similar to the learning goal effect, there is some evidence in the literature of, categorisation using rules or, categorisation using instances, occurring on the same material depending on the instructions for the task.

Take, for instance, a study by Rips (1989). Rips showed that if subjects were asked to say whether an item is more '*likely to be*' object X or object Y then they use rules to categorise the object. For example, "is a 3-in. circular item more *likely to be* a pizza or a quarter" resulted in most subjects saying pizza as they were applying a rule to categorise the item (such as, if its diameter is not close to ¾ of an inch then it's not a quarter). If however subjects were asked to say whether an item is more '*similar*' to object X or object Y then subjects use similarity judgements - that is, taking the instance of an object and comparing it to instances of objects X & Y. In the case of the above example, most subjects categorised the 3-in. circular item as a quarter as it was much closer in size to a quarter than to the average sized pizza. Similarly, in a follow up study to Rips's, Smith and Sloman (1994) showed that categorisation of the same item using rules or categorisation using instances depended on how much information was given about the item to be categorised. Sparse descriptions of the items combined with 'think aloud' instructions resulted in categorisation by using rules whereas rich descriptions of the items resulted in categorisation by using instances. Therefore in the concept learning literature, like in this thesis, there is evidence for both rule learning and instance learning occurring on the same task as a function of the experiment's instructions.

As mentioned in the paragraph before last, most of the concept learning models focus on either explaining rule learning *or* instance learning, not both. However in the light of some of the work suggesting that both can occur on the same material, there are some models like the dual space model which offer explanations for the two (e.g. Wisniewski and Medin's (1995) interactive model, Michalski's (1989) two-tier approach). As discussed in the discussion of Experiment 3 (see page 99), the interactive model only attempts to explain learning in relation to how much prior knowledge subjects have and can not explain the learning goal effect shown in this thesis. Briefly, the two-tier model suggests that learning is based on instances from long term memory and general rules applied to those instances and new instances to help categorise the new instances. There is no attempt in the model to explain when just instances or rules will be used. Hence the two-tier model also cannot be used to explain the learning goal effect.

So, the distinction between instances and rule learning is well documented in the concept learning literature, although, very few models try to account for both kinds of learning. Those that do cannot explain the learning goal effect. Only the dual space model seems able to do this as only the dual space model provides a framework, a 'problem space', where both kinds of learning can occur depending on how the situation encourages subjects to explore one part of problem space or another - i.e. explore instance space only or rule space and instance space. It would be interesting to see if concept learning could also be affected by the learning goal in the same manner as has been shown in the experiments in this thesis. It would also be interesting to see if the dual space model could be used as a basis for providing a more integrated account of the two kinds of concept learning.

One last point to make about the dual space model is that it requires two theoretical spaces, instance and rule space, within the problem space. Both of these two spaces require their own set of theories explaining the mechanics of learning. One possibility that further work may raise is whether the need for rule space is necessary. If further work on the instance learning of the pattern search subjects concludes that instance learning occurs for the pattern search subjects and is in fact vital for their rule inductions then the notion of rule space may not be needed. However, as described above on the section about the pattern search learners, the present set of results suggest that instance learning for pattern search subjects does not play a crucial part in their rule learning abilities. Therefore it is likely that rule space will remain an essential part of any model explaining the learning goal effect.

Models of implicit instance learning

This is essentially a recap of points made in the discussion of the random number generation experiment, (see page 124). The question is, how exactly should the implicit instance learning, as shown in this thesis, be modelled? It was proposed in the random number generation experiment that there are two possible models that could be used to describe it - Anderson's (1987) ACT model and Logan's (1988) instance theory model. Anderson's ACT model can explain implicit instance learning as it allows for actions to be outside working memory. Once certain steps of action have been repeatedly carried out using working memory they become proceduralised. Then, the procedure itself is automatically executed given the appropriate stimulus. The ACT model can account for the successful <u>use</u> of implicit instance learning, but not the learning itself. The appropriate stimulus could be the match between a task situation and a memorised instance and the proceduralised steps of action could be to perform the prescribed action stored with the instance given that it was successful on a previous occasion. However the model cannot explain how instances are learned and proceduralised in the first place. The model dictates that the steps taken to use instances should pass through working memory repeatedly until they become proceduralised. The results of the random number generation experiment suggest that subjects are just as capable of performing implicit instance learning when their working memory is occupied by a neutral task *right from the beginning* of the learning episode. One possible way to model implicit instance learning using the ACT model is to enhance it by providing a mechanism where some steps of actions are proceduralised automatically. Considering the usefulness of the ACT model in explaining other areas of learning, this may be an appropriate endeavour as it would then allow the model to explain both implicit and explicit learning.

For the sake of just trying to explain implicit instance learning this is not necessary as another model, Logan's instance theory, can explain implicit instance learning without the need for it to be adjusted. Logan's instance theory has already been re-reviewed in this Discussion (see pg. 204). Briefly summarising, the model assumes that people accumulate instances throughout their learning, however to begin with they use general strategies to perform a task and only when enough instances have been accumulated, does instance learning dominate. The key advantage over the ACT model is that Logan's theory does not assume that either the acquisition or the use of instances has to occur in working memory at any point. Therefore, contrary to the ACT model, Logan's model can explain both the use of implicit instances and the implicit acquisition of them in the first place. This model also coherently explains the apparent implicit and explicit mix of instance learning that control task subjects exhibit, (again see pg. 204). So, all in all, the favoured model to explain implicit instance learning as shown in this thesis, is Logan's instance theory.

Reber's evolutionary model of learning

Arthur Reber has put forward the proposition that neurological systems underpinning implicit learning processes should be older from an evolutionary perspective than explicit systems. His suggestion stems from his Axiom About Consciousness - "Consciousness is a late arrival on the evolutionary scene. Sophisticated unconscious perceptual and cognitive functions preceded its emergence by a considerable margin (Reber, 1993, pg. 86)". There are a number of predictions that can be made from such a claim. These concern the characteristics of implicit learning modes relative to explicit ones (e.g. Robustness - implicit learning modes should be robust in the face of dysfunctions that affect explicit learning modes; Age independence - they should be less affected by age and developmental level than explicit learning processes; Low Variability - described below; IQ independence - there should be less correlation with IQ than explicit processes; Commonality of process - the processes of implicit learning should have more in common with learning systems of other species). The predicted characteristic that is easily testable in the experiments of this thesis is the one of Low Variability. The idea is, that since the neurological 'hardware' underpinning implicit learning is supposedly older than that underpinning explicit learning, then it has had more time to be refined through evolutionary processes. Therefore, the end result is more reliable and stable and so replicated more precisely through individuals than the newer explicit neurological mechanisms. This leads to the prediction that "Population variances should be much smaller when implicit processes are measured than when explicit processes are measured (pg. 88)".

To test such a prediction the variances of the measures used for the different goal groups can be examined. The prediction questionnaire is the measure where the particular modes of learning are most easily distinguishable as each mode of learning has its own unique pattern of prediction question performance. Therefore, to test the idea that implicit learners' performance should show lower variance than explicit learners' performance, the variance of the total prediction question scores shall be examined. The average variance for the total prediction question scores of the groups that were deemed to be learning purely explicitly (the rule learners) was 683.25[6], and for the groups that were deemed to be learning purely implicitly (the implicit instance learners) the average variance was 308.72[7]. Admittedly, grouping data across all the experiments with their various differences is a fairly crude method, however as can be seen, in line with Reber's prediction, the variance for the explicit learners is over twice that of implicit learners. This is not quite as large a difference as that observed by Reber, Walkenfeld and Hernstadt (1991) who showed explicit learners having 4 times the variance of implicit learners. Reber, however, suggests that the variance of implicit learning can always be increased as there are frequently explicit processes used in the measure that examines implicit

---

[6] The explicit rule learning groups were: the pattern search subjects from the learning goal effect experiment and the transfer experiments, the pattern search models and pattern search observers from the observing experiment, and, the explaining pattern search subjects from the concurrent verbalisation experiment.

[7] The implicit instance learning groups were: the dual goal subjects from the learning goal effect experiment, all the subjects from the random number generation experiment and the subjects in the two dual goal groups from the concurrent verbalisation experiment.

learners performance (Reber, 1993, pg. 100). Still, the crude results obtained from the thesis as a whole are in the same direction of Reber's prediction.

The other interesting result is the variance for the groups learning through a mixture of implicit and explicit processes, that is the subjects with a control task goal. The average variance for the total prediction question scores for these subjects was a value of 278.76[8]. Clearly this value is more in line with the magnitude of the implicit instance learners' variance than the explicit rule learners'. This is interesting because, if Reber's line of thinking is to be taken seriously, it implies that control task subjects' learning processes are underpinned by evolutionarily older mechanisms than those used by the pattern search subjects. This is understandable as essentially the results have suggested that both dual goal and control task subjects are doing the same thing - instance learning, it is just that the dual goal subjects do not need explicit processes. It is possible, therefore, that it is instance learning whether implicit or explicit, rather than specifically implicit learning that is the evolutionarily older mechanism.

Methodological Issues And Implications

A potential flaw and confounding factor in the design of the prediction questions

In the discussion of the memory experiment it was shown that subjects with a dual goal were recognising Old-wrong instances above chance when it had been predicted that they should perform at chance. One of the reasons put forward for this was that subjects may not actually have a memory of Old-wrong instances (created from the test phase), but, may have a memory of Old-correct instances from the learning phase that were identical to the Old-wrong instances from the test phase. This would mean that subjects' performance on the Old-wrong instances was not a reflection of how well they remembered situations they performed wrongly in. (Incidentally, the analysis was redone with the conflicting data removed and an identical pattern of results was still shown.)

[8] The groups used to generate this value were all the groups with a control task learning goal excluding the control task instructioned observers from the observation experiment. Their particular experimental condition meant that they learned nothing. Also excluded, like in the other two calculations, were the groups from the memory experiment as they did not have a prediction questionnaire.

This line of thinking can also be applied to the prediction questions in every other experiment of the thesis (apart from Experiment 2 as this has its prediction questions generated from the learning phase). It is possible that prediction questions that are asking subjects to make predictions from Old-wrong situations (generated from test phase trials) may actually represent identical situations that in the learning phase were Old-correct situations. The opposite is also true - Old-correct prediction situations could actually be Old-wrong prediction situations in terms of the learning phase. This could confound the validity of the results as it would violate the integrity of the prediction question type. Considering that the memory experiment showed that it did not affect the results it is probably safe to assume that it will not affect the other results. However, to be certain, all the results relating to the prediction questions were reanalysed with the conflicting data between learning and test phases removed. As for the data reported through out the thesis, the reanalyses also removed the situations where, within the test phase, the Old-wrong and Old-correct situations were identical (e.g. see pg. 28). The results of these reanalyses can be seen in the Appendix with the F tables and other additional statistics connected to each experiment (for the learning goal effect experiment the reanalysis can be seen in Appendix 1, pg. 231, for the concurrent verbalisation experiment, Appendix 3, pg. 241, for the random number generation experiment, Appendix 4, pg. 246, for the transfer experiments, Appendix 6, pg. 253 & 257). For every experiment, the result shows an identical pattern of results to that shown for the original analysis reported in the experiments, with all the results that were significant before remaining significant (all p values < 0.05). Hence, this potential confounding factor does not alter the findings of the thesis in any way.

The prediction questions as a tool for assessing learning.

The importance of designing the prediction questions in the careful manner used throughout this thesis has been repeatedly described (for example, in this Discussion see pages, 205, 207). What is discussed here is exactly how prediction questions have been used in this thesis and in the past to provide evidence on the nature of subjects' learning. The common assumption is that prediction questions - the arbitrarily created old style prediction questions -

tap explicit knowledge (e.g. Berry and Broadbent, 1984; 1988; Shanks and St. John, 1994). Indeed performance at chance on prediction questions was taken as a sign of lack of explicit knowledge. The point being made in this paragraph, is that though prediction questions can be used as a measure of explicit knowledge, they can also be used to measure aspects of implicit knowledge. Designed with a distinct set of Old-wrong, Old-correct and New prediction questions, they can be used to demonstrate explicit rule learning, explicit instance learning and implicit instance learning. What is important is the *pattern* of performance across these prediction question types. Good comparable performance on all prediction questions, particularly New questions, is a sign of rule learning. Good performance only on Old-wrong and Old-correct questions is a sign of instance learning guided by a look-up table consisting of all instances encountered. Good performance only on Old-correct instances, according to the conclusions made in the random number generation experiment, reflects pure implicit instance learning.

Practical Issues And Implications

What practical implications do the results have for training courses and problem setting strategies? As the focus of the thesis has been on the learning goal effect, the implications of the results are towards which learning goal is the best to set subjects during a training programme or on a set of training problems. The results indicate that the pattern search goal is the best to set students. Subjects with this goal are significantly better at controlling the system, are better at making predictions for the system, and have more accurate and accessible knowledge of the mechanics of the system than subjects with the other goals. The results also suggest that pattern search learners, on top of their knowledge of rules, appear to acquire knowledge of instances like the other goal groups. Therefore, there is no point in setting subjects, say, a control task goal so that they can memorise instances and achieve automaticity, because subjects will acquire the instances from the pattern search goal anyway and probably explicit rules as well. (As mentioned above, this last point needs some extra work to confirm the finding as it may be that when pattern search subjects have no control goal at any point in their training then they may not learn instances.) One final bonus of setting a pattern search goal is that the resulting learning transfers to a novel, but structurally identical task, which will not occur with the learning resulting from the other goals. In addition to setting a pattern search goal (or indeed any other goal) the results also suggest that students should be encouraged to make self-explanations as this should considerably enhance learning.

It may be that it is not possible to set a pattern search goal in some training conditions and a control task goal is the only option. If this is the case than it is important to keep the cognitive load light, otherwise the only learning that may be possible will be the implicit instance learning of the dual goal group. Though robust in the face of a heavy cognitive load, implicit instance learning is not as productive as control task instance learning. For example, there is some evidence that control task instance learners are better at controlling the person interaction task (as shown by the learning goal effect experiment), however this does not always happen (as shown by the memory experiment). What is certain is that dual goal subjects (implicit instance learners) are worse at making predictions from Old-wrong situations than they from Old-correct situations. This is not the case for the control task instance learners. It is obviously sensible to set the goal condition where the most information will be learnt. Therefore, if the pattern search goal is not an option then it is advisable to set the control task goal, with every effort made to reduce cognitive load.

What can the thesis say about the potential advantages to learners of varying the specificity of a learning goal? No firm advice on this issue can come from this thesis as it has really been examining completely different learning goals (both specific in their own manner) not specificity of the goal per se. Still though, it may be possible to take some advice from the work of the thesis regarding this matter. Consider the following: The important impact goal specificity should have on learning is by adjusting the area of problem space that a subject will examine. A completely specific goal, such as the control task goal, will minimise the area of problem space that will be examined. On the other hand, a minimally specific goal will maximise the area of problem space that will be examined. The result of this minimally specific goal is similar to what occurs with the pattern search goal. The pattern search goal encourages subjects to explore as much of problem space (instance and rule space) as they wish. Therefore, the pattern search goal could almost be considered to be examining the effect of a goal with the minimum specificity possible. In that case, considering the control task goal is examining the effect of a completely specific goal, the results of the thesis can give some advice as to how specific to make a goal.

It may be best to recommend setting goals in training situations with the minimal specificity possible as then explicit rule learning (with all its advantages) may be more likely. However the comparison between a goal of minimal specificity and a pattern search goal may not be appropriate as the pattern search goal is specific in its own manner in that it gives subjects a specific purpose. Therefore, it would be sensible to examine the effect of goal specificity per se. The full range of specificities should be examined because even if this thesis can give information on the effect of goal specificity, it certainly has not explored the effect of setting a goal with middling specificity or any other specificity in between the two extremes of completely and minimally specific. It is important to study the effect of goal specificity as it may be that in a particular training situation a pattern search goal cannot be set, though a different goal with low specificity may be a possible option. However, there would be no point in doing so until further work had shown that it was in the learners' best interests.

One final issue that needs to be discussed is the appropriateness of taking advice from the work in this thesis considering the laboratory setting of the work. In other words, has the learning studied been ecologically valid enough to make recommendations to training programmes? On the whole, I would like to argue yes, however only time can tell. Some of the advice mentioned above would be difficult to apply to some learning situations. For instance, if you have students learning say a new computer programming language, what is the best way to teach them or to set them learning problems considering the results of this thesis? There are a number of approaches to determining how valid the results are to the 'real world'. One is to test it experimentally - try and set the different goals used in the thesis and examine the results. Another, (and one way of exploring the last suggestion), is to arrange training situations to be as similar to a dynamic system (the learning situations studied in the thesis) as possible. For instance, in the area of trigonometry a computer programme could be set up to endlessly create trigonometry problems[9]. The rules of trigonometry would essentially be the rules underlying the system. Students could then practise the theory they have been taught on these problem sets or even try to discover it for themselves. By altering training programmes to be more similar to dynamic systems, it would mean that the implications and advice from the work of this thesis would be more likely to have a useful impact. On one last note, I feel that the advice from the thesis is likely to have some 'real world' application as the whole work in the thesis stemmed from Owen and Sweller's (1985) 'real world' study.

Summary of the Discussion

The discussion started off by briefly reviewing the conclusions from the six experimental chapters that make up the main body of the thesis. The theoretical implications and issues relating to the results of each goal group were then discussed.

As regards theoretical issues for the control task goals, the results show that the control task goal causes instance learning due to a direct influence on cognitive processes, and that it is the control task goal not the salience of the person interaction task that is the more fundamental variable causing instance learning. Further work should explore how the variables of learning goal and salience interact as this may be important when regarding the practical implications of the learning goal effect. The results show that control task instance learners are learning through a mixture of implicit and explicit processes. This finding was only possible due to the careful manner in which the prediction questions were constructed. The notion stemming from the memory experiment that instances are encoded in some abstract manner is probably not true. Tentative support for this was taken from the fact that the transfer experiments did not show far transfer for the control task subjects, something that would be more likely if their learning consisted of instances encoded in an abstract manner.

---

[9] For instance, the programme could be constructed to produce triangles with both givens (angle sizes, lengths of sides) and unknowns. The students could calculate the unknowns and enter them into the triangles. If they were correct then the triangle would remain the same and the next triangle problem would appear. However, if they were wrong then the triangle and the givens would change to what they should be for the unknown to be what the subjects had suggested.

The results of the dual goal subjects were then discussed. These subjects appear to be learning purely implicitly. The evidence provided in the thesis for this claim has a number of advantages over other studies. These include better measures and procedures to rule out any chance of subjects being aware during learning, and the demonstration of qualitative differences between learning modes. Assuming that action is vital to implicit learning not explicit learning, then using the observation paradigm with the dual goal subjects may provide evidence of a double dissociation between implicit and explicit modes of learning, thus giving even better evidence of implicit learning. However, feedback not action may be the vital requirement for implicit learning. In a recap of the dual goal subjects' results for the memory experiment, it was also suggested that the experiment could be re-run with confidence rating questions added to each recognition judgement, thus then being able to determine if dual goal subjects do have weaker memories for Old-wrong instances.

The pattern search subjects results were then discussed. It was suggested that subjects' apparent learning of instances as well as rules may not occur if the memory experiment was redesigned with the memory test placed before a specific control goal is given to subjects. Also, the apparent use of instances over rules in recognition may not be seen if the experiment was redesigned with more information in the instances, thus making the application of rules easier. Instance learning is probably not vital to the pattern search subjects' rule learning as their rule inducing was unaffected by a reduction in the range of instances they saw during the observation experiment. Hypothesis testing not instance learning is probably the primary cause of rule learning, because when hypothesis testing is interfered with (as in the concurrent verbalisation experiment), subjects' rule learning suffers. A further study collecting undirected (i.e. unlike the concurrent verbalisation experiment) verbal protocols would be useful in getting a firmer grasp of the exact processes that subjects are using to learn.

The success of the dual space model as an explanatory tool to explain the learning goal effect was then summarised. In the unlikely event that future work showed that instance learning was a vital part of the pattern search subjects' rule learning then the dual space model may be inappropriate as the notion of rule space would no longer be needed to explain the effect of the pattern search instructions. However, given the present results, it was argued that only the dual space model can explain the learning goal effect, an effect that cannot be explained by the combined instance and rule learning models in the concept learning literature.

A prediction from Arthur Reber was then tested. Reber has suggested that explicit rule learners should have a higher variance on measures of learning than implicit learners. This prediction was shown to hold true for the results of the thesis. The mean variance for the total prediction scores for all groups in the thesis considered to be learning rules explicitly was double that of groups thought to be learning instances purely implicitly. However, control task subjects had a mean variance of total prediction question scores in the same order as that of the dual goal instance learners. Consequently, it is possible that the critical distinction is between instance learning and rule learning rather than between implicit and explicit learning.

Some methodological issues of the thesis were then considered focusing on the prediction questions. One possible flaw in the way they were designed was described, and reanalyses were done exploring the results with the flaw removed. It was shown that with the potential flaw removed, the pattern of results was identical to that shown in the rest of the thesis. It was also pointed out that prediction questions, though previously used to tap explicit processes, if designed carefully can be used to reveal the nature of learning, be it explicit rule learning, explicit instance learning or implicit instance learning.

Finally the practical implications of the thesis were considered. If at all possible a pattern search goal should be set to subjects as this is the most productive goal. Otherwise, a control task goal should be set, with care being taken to reduce the cognitive load. Further work could be done on the effect on learning of varying goal specificity systematically as it has not been studied per se in this thesis. Finally, the appropriateness of making practical advice from the results of the thesis were considered in light of the lab based work of the experiments. It was suggested that further experiments could be carried out to explore how ecologically valid the results are. Alternatively, 'real world' training programmes could be designed so that they represent a dynamic system thus increasing the chances of advice from the thesis being applicable.

APPENDIX 1
EXPERIMENT 1'S F TABLES AND ADDITIONAL STATISTICS

Learning phase data

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	284.25	44	6.46		
INST_ORD	13.50	1	13.50	2.09	.155
LRN GOAL	160.17	1	160.17	24.79	.000
INST_ORD BY LRN GOAL	2.04	1	2.04	.32	.577

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	260.08	44	5.91		
TRL BLOCK	26.04	1	26.04	4.41	.042
INST_ORD BY TRL BLOCK	.17	1	.17	.03	.867
LRN GOAL BY TRL BLOCK	10.67	1	10.67	1.80	.186
INST_ORD BY LRN GOAL BY TRL BLOCK	12.04	1	12.04	2.04	.161

Test phase data

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	681.00	66	10.32		
INST_ORD	3.36	1	3.36	.33	.570
LRN GOAL	401.01	2	200.51	19.43	.000
INST_ORD BY LRN GOAL	40.60	2	20.30	1.97	.148

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	465.83	66	7.06		
TRL BLOCK	2.78	1	2.78	.39	.533
INST_ORD BY TRL BLOCK	.00	1	.00	.00	1.000
LRN GOAL BY TRL BLOCK	18.10	2	9.05	1.28	.284
INST_ORD BY LRN GOAL BY TRL BLOCK	7.29	2	3.65	.52	.599

Between Groups comparisons for total test phase score and first and last 15 trials score

non-specific goal group vs specific goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	357.52083	1151.95833	357.52083	25.04257	14.27652	.000
F15_TST	56.33333	405.33333	56.33333	8.81159	6.39309	.015
L15_TST	130.02083	448.29167	130.02083	9.74547	13.34167	.001

non-specific goal group vs dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	768.00000	901.91667	768.00000	19.60688	39.16992	.000
F15_TST	221.02083	347.95833	221.02083	7.56431	29.21890	.000
L15_TST	165.02083	413.45833	165.02083	8.98822	18.35967	.000

specific goal group vs dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	77.52083	845.95833	77.52083	18.39040	4.21529	.046
F15_TST	54.18750	313.29167	54.18750	6.81069	7.95624	.007
L15_TST	2.08333	467.83333	2.08333	10.17029	.20485	.653

EXPERIMENT 1'S F TABLES AND ADDITIONAL STATISTICS

Transfer between learning and test phases

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	350.96	44	7.98		
LRN GOAL	173.34	1	173.34	21.73	.000
INST_ORD	1.76	1	1.76	.22	.641
LRN GOAL BY INST_ORD	.09	1	.09	.01	.914

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	330.62	44	7.51		
TRL BLOCK	.51	1	.51	.07	.796
LRN GOAL BY TRL BLOCK	7.59	1	7.59	1.01	.320
INST_ORD BY TRL BLOCK	3.76	1	3.76	.50	.483
LRN GOAL BY INST_ORD BY TRL BLOCK	3.01	1	3.01	.40	.530

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	775.63	44	17.63		
LRN GOAL	356.51	1	356.51	20.22	.000
INST_ORD	2.34	1	2.34	.13	.717
LRN GOAL BY INST_ORD	5.51	1	5.51	.31	.579

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	627.96	44	14.27		
PHASE SCORE	21.09	1	21.09	1.48	.231
LRN GOAL BY PHASE SCORE	41.34	1	41.34	2.90	.096
INST_ORD BY PHASE SCORE	33.84	1	33.84	2.37	.131
LRN GOALBY INST_ORD BY PHASE SCORE	.26	1	.26	.02	.893

Prediction questions - Analysis done on Figure 1.3

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	95880.32	66	1452.73		
INST_ORD	2523.78	1	2523.78	1.74	.192
LRN GOAL	52438.27	2	26219.14	18.05	.000
INST_ORD BY LRN GOAL	2950.10	2	1475.05	1.02	.368

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	70627.31	132	535.06		
QUE_TYPE	4861.19	2	2430.59	4.54	.012
INST_ORD BY QUE_TYPE	509.59	2	254.80	.48	.622
LRN GOAL BY QUE_TYPE	4308.18	4	1077.04	2.01	.096
INST_ORD BY LRN GOAL BY QUE_TYPE	656.69	4	164.17	.31	.873

Between Groups comparisons for total prediction questionnaire scores

non-specific goal group vs specific goal group, then
non-specific goal group vs dual goal group, then
specific goal group vs dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
PQ_TOT	11988.2430	28271.4662	11988.2430	614.59709	19.50586	.000
PQ_TOT	16374.9517	28701.0154	16374.9517	623.93512	26.24464	.000
PQ_TOT	341.28183	12388.4578	341.28183	269.31430	1.26723	.266

EXPERIMENT 1'S F TABLES AND ADDITIONAL STATISTICS

Prediction questions - Analysis done when prediction questions defined by the last two elements rather than all three.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	94380.03	65	1452.00		
INST_ORD	757.51	1	757.51	.52	.473
LRN_GOAL	47269.30	2	23634.65	16.28	.000
INST_ORD BY LRN_GOAL	1426.42	2	713.21	.49	.614

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	82542.87	130	634.95		
QUE_TYPE	6167.23	2	3083.62	4.86	.009
INST_ORD BY QUE_TYPE	1683.17	2	841.58	1.33	.269
LRN_GOAL BY QUE_TYPE	5586.67	4	1396.67	2.20	.073
INST_ORD BY LRN_GOAL BY QUE_TYPE	2087.77	4	521.94	.82	.513

Between Groups comparisons for total prediction questionnaire scores

non-specific goal group vs specific goal group,
non-specific goal group vs dual goal group, then
specific goal group vs dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
PQ_TOT2	10984.1163	26627.2969	10984.1163	591.71771	18.56310	.000
PQ_TOT2	12465.7607	25888.3954	12465.7607	575.29768	21.66837	.000
PQ_TOT2	47.87310	11835.0697	47.87310	257.28412	.18607	.668

Within group comparisons

Non-specific goal group

OC vs New	Z = -.5241	2-Tailed P = .6002
Ow vs New	Z = -.2272	2-Tailed P = .8203
Ow vs Oc	Z = -.1177	2-Tailed P = .9063

specific goal group

OC vs New	Z = -2.3700	2-Tailed P = .0178
Ow vs New	Z = -2.9718	2-Tailed P = .0030
Ow vs Oc	Z = -.1867	2-Tailed P = .8519

dual goal group

OC vs New	Z = -2.2586	2-Tailed P = .0239
Ow vs New	Z = -.0747	2-Tailed P = .9405
Ow vs Oc	Z = -2.2864	2-Tailed P = .0222

Between Groups comparisons for each prediction question type

non-specific goal group vs specific goal group

	U	W	Z	2-Tailed P
OW	157.5	457.5	-2.7218	.0065
OC	149.0	679.0	-2.7294	.0063
NEW	104.5	404.5	-3.8334	.0001

non-specific goal group vs dual goal group

	U	W	Z	2-Tailed P
OW	116.0	760.0	-3.5922	.0003
OC	186.0	642.0	-1.9510	.0511
NEW	114.0	762.0	-3.6557	.0003

specific goal group vs dual goal group

	U	W	Z	2-Tailed P
OW	178.0	698.0	-2.3038	.0212
OC	256.0	556.0	-.6643	.5065
NEW	287.5	588.5	-.0106	.9916

EXPERIMENT 1'S F TABLES AND ADDITIONAL STATISTICS

Spearman Rank correlations between total prediction question scores with total learning phase and test phase scores

	specific goal group	Nsgg	dual goal group
TOT_TST	.0562	.7331	.4784
	N(24)	N(23)	N(24)
	Sig .794	Sig .000	Sig .018
TOT_LRN	.0961		.0129
	N(24)		N(24)
	Sig .655		Sig .952
	TOT_QU2	TOT_QU2	TOT_QU2

Prediction questions - Analysis done when prediction questions have Old-wrong and Old-correct questions refined so that there are no Old-wrong situations that were Old-correct situations in the learning phase and also so that there were no Old-correct situations that were Old-wrong situations in the learning phase.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	99734.24	64	1558.35		
INST_ORD	2700.24	1	2700.24	1.73	.193
LRN_GOAL	49399.03	2	24699.52	15.85	.000
INST_ORD BY LRN_GOAL	3211.13	2	1605.56	1.03	.363

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	65877.75	128	514.67		
QUE_TYPE	4108.27	2	2054.14	3.99	.021
INST_ORD BY QUE_TYPE	320.82	2	160.41	.31	.733
LRN_GOAL BY QUE_TYPE	5450.96	4	1362.74	2.65	.036
INST_ORD BY LRN_GOAL BY QUE_TYPE	852.46	4	213.12	.41	.798

Between Groups comparisons for total prediction questionnaire scores

non-specific goal group vs specific goal group,
non-specific goal group vs dual goal group, then
specific goal group vs dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
PQ_TOT3	9487.50000	28554.9383	9487.50000	648.97587	14.61919	.000
PQ_TOT3	14060.8696	28130.0926	14060.8696	639.32029	21.99347	.000
PQ_TOT3	468.75000	13595.2160	468.75000	295.54817	1.58604	.214

Within group comparisons

Non-specific goal group

OC vs New	Z = -.1177	2-Tailed P = .9063
Ow vs New	Z = -1.0590	2-Tailed P = .2896
Ow vs Oc	Z = -1.2944	2-Tailed P = .1955

specific goal group

OC vs New	Z = -2.2077	2-Tailed P = .0273
Ow vs New	Z = -2.3288	2-Tailed P = .0199
Ow vs Oc	Z = -.0974	2-Tailed P = .9224

dual goal group

OC vs New	Z = -2.3051	2-Tailed P = .0212
Ow vs New	Z = -.0947	2-Tailed P = .9246
Ow vs Oc	Z = -2.2726	2-Tailed P = .0231

Between Groups comparisons for each prediction question type

non-specific goal group vs specific goal group

	U	W	Z	2-Tailed P
OW	128.0	700.0	-3.1922	.0014
OC	162.0	666.0	-2.4516	.0142

non-specific goal group vs dual goal group

	U	W	Z	2-Tailed P
OW	83.0	745.0	-4.1669	.0000
OC	177.5	650.5	-2.1261	.0335

specific goal group vs dual goal group

	U	W	Z	2-Tailed P
OW	169.0	707.0	-2.4921	.0127
OC	285.0	585.0	-.0626	.9501

Spearman Rank correlations between total prediction question scores with total learning phase and test phase scores

	specific goal group	Nsgg	dual goal group
TOT_TST	.2120	.8617	.3852
	N(24)	N(22)	N(24)
	Sig .320	Sig .000	Sig .063
TOT_LRN	-.0510		.0619
	N(24)		N(24)
	Sig .813		Sig .774
	TOT_QU3	PTOT_QU3	TOT_QU3

EXPERIMENT 2'S F TABLES AND ADDITIONAL STATISTICS

MODELS' DATA

test phase

Comparisons to chance (a value of 7.4) for total test phase score

pattern search models	$t(11) = 2.56$, $p = 0.027$
control task models	$t(11) = 3.4$, $p = 0.006$

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	484.83	22	22.04		
LRN GOAL	.08	1	.08	.00	.952

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	158.67	22	7.21		
TRL BLOCK	33.33	1	33.33	4.62	.043
LRN GOAL BY TRL BLOCK	27.00	1	27.00	3.74	.066

prediction questions

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	19110.02	22	868.64		
LRN GOAL	16729.48	1	16729.48	19.26	.000

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	6719.28	22	305.42		
QUE_TYPE	3070.22	1	3070.22	10.05	.004
LRN GOAL BY QUE_TYPE	1450.78	1	1450.78	4.75	.040

Between group comparisons for the Old and New questions (models: control task vs pattern search)

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
OLD	4163.59311	15395.9619	4163.59311	699.81645	5.94955	.023
NEW	14016.6667	10433.3333	14016.6667	474.24242	29.55591	.000

Comparisons to chance (a value of 24%) of total prediction questionnaire scores and Old and New scores

pattern search models

total prediction question scores	$t(11) = -6.06$, $p = .000$
Old	$t(11) = -6.21$, $p = .000$
New	$t(11) = -5.33$, $p = .000$

control task models

total prediction question scores	$t(11) = -1.17$, $p = .266$
Old	$t(11) = -2.80$, $p = .017$
New	$t(11) = 1.47$, $p = .170$

APPENDIX 2
EXPERIMENT 2'S F TABLES AND ADDITIONAL STATISTICS

OBSERVERS' DATA

comparisons to chance

	total test phase score	New	Old
CTCT	t(11) = -0.11, p = 0.912	t(11) = -0.05, p = 0.964	t(11) = -0.43, p = 0.677
PSCT	t(11) = -1.07, p = 0.308	t(11) = 0.64, p = 0.537	t(11) = -0.04, p = 0.969
CTPS	t(11) = 3.30, p = 0.007	t(11) = -2.90, p = 0.014	t(11) = -3.74, p = 0.003
PSPS	t(11) = 4.38, p = 0.001	t(11) = -6.10, p = 0.000	t(11) = -7.63, p = 0.000

test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	840.29	44	19.10		
GL_OF_MO	.51	1	.51	.03	.871
GL_OF_OB	446.34	1	446.34	23.37	.000
GL_OF_MO BY GL_OF_OB	25.01	1	25.01	1.31	.259

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	184.12	44	4.18		
TRL BLOCK	23.01	1	23.01	5.50	.024
GL_OF_MO BY TRL BLOCK	4.59	1	4.59	1.10	.300
GL_OF_OB BY TRL BLOCK	.26	1	.26	.06	.804
GL_OF_MO BY GL_OF_OB BY TRL BLOCK	17.51	1	17.51	4.18	.047

prediction questions

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	37758.33	44	858.14		
GL_OF_MO	1204.17	1	1204.17	1.40	.243
GL_OF_OB	40837.50	1	40837.50	47.59	.000
GL_OF_MO BY GL_OF_OB	1350.00	1	1350.00	1.57	.216

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	8975.00	44	203.98		
QUE_TYPE	266.67	1	266.67	1.31	.259
GL_OF_MO BY QUE_TYPE	37.50	1	37.50	.18	.670
GL_OF_OB BY QUE_TYPE	204.17	1	204.17	1.00	.323
GL_OF_MO BY GL_OF_OB BY QUE_TYPE	16.67	1	16.67	.08	.776

Between group comparisons for the New and Old questions between all sets of subjects.

Column 1 of Table 2.3

Control task models vs pattern search models

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
OLD	4163.59311	15395.9619	4163.59311	699.81645	5.94955	.023
NEW	14016.6667	10433.3333	14016.6667	474.24242	29.55591	.000

Control task models vs CTPS observers

OLD	1292.29681	19829.2953	1292.29681	901.33160	1.43376	.244
NEW	7704.16667	16258.3333	7704.16667	739.01515	10.42491	.004

Control task models vs PSPS observers

OLD	5283.96348	13529.2953	5283.96348	614.96797	8.59226	.008
NEW	15000.0000	8933.33333	15000.0000	406.06061	36.94030	.000

Control task models vs CTCT observers

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
OLD	3125.63014	9254.29527	3125.63014	420.64978	7.43048	.012
NEW	204.16667	3658.33333	204.16667	166.28788	1.22779	.280

Control task models vs PSCT observers

OLD	2685.81533	9920.96193	2685.81533	450.95282	5.95587	.023
NEW	66.66667	3733.33333	66.66667	169.69697	.39286	.537

Column 2 of Table 2.3

Pattern search observers vs CTPS observers

OLD	816.66667	19966.6667	816.66667	907.57576	.89983	.353
NEW	937.50000	22758.3333	937.50000	1034.46970	.90626	.351

Pattern search observers vs PSPS observers

OLD	66.66667	13666.6667	66.66667	621.21212	.10732	.746
NEW	16.66667	15433.3333	16.66667	701.51515	.02376	.879

Pattern search observers vs CTCT observers

OLD	14504.1667	9391.66667	14504.1667	426.89394	33.97604	.000
NEW	10837.5000	10158.3333	10837.5000	461.74242	23.47088	.000

Pattern search observers vs PSCT observers

OLD	13537.5000	10058.3333	13537.5000	457.19697	29.60978	.000
NEW	12150.0000	10233.3333	12150.0000	465.15152	26.12052	.000

Column 3 of Table 2.3

CTPS observers vs PSPS observers

OLD	1350.00000	18100.0000	1350.00000	822.72727	1.64088	.214
NEW	1204.16667	21258.3333	1204.16667	966.28788	1.24618	.276

CTPS observers vs CTCT observers

OLD	8437.50000	13825.0000	8437.50000	628.40909	13.42676	.001
NEW	5400.00000	15983.3333	5400.00000	726.51515	7.43274	.012

CTPS observers vs PSCT observers

OLD	7704.16667	14491.6667	7704.16667	658.71212	11.69580	.002
NEW	6337.50000	16058.3333	6337.50000	729.92424	8.68241	.007

Column 4 of Table 2.3

PSPS observers vs CTCT observers

OLD	16537.5000	7525.00000	16537.5000	342.04545	48.34884	.000
NEW	11704.1667	8658.33333	11704.1667	393.56061	29.73917	.000

PSPS observers vs PSCT observers

OLD	15504.1667	8191.66667	15504.1667	372.34848	41.63886	.000
NEW	13066.6667	8733.33333	13066.6667	396.96970	32.91603	.000

Column 5 of Table 2.3

CTCT observers vs PSCT observers

OLD	16.66667	3916.66667	16.66667	178.03030	.09362	.763
NEW	37.50000	3458.33333	37.50000	157.19697	.23855	.630

APPENDIX 3
EXPERIMENT 3'S F TABLES AND ADDITIONAL STATISTICS

learning trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	651.50	44	14.81		
LRN GOAL	26.04	1	26.04	1.76	.192
VERB	.04	1	.04	.00	.958
LRN GOAL BY VERB	9.38	1	9.38	.63	.430

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	217.50	44	4.94		
TRL BLOCK	10.67	1	10.67	2.16	.149
LRN GOAL BY TRL BLOCK	2.67	1	2.67	.54	.467
VERB BY TRL BLOCK	24.00	1	24.00	4.86	.033
LRN GOAL BY VERB BY TRL BLOCK	.17	1	.17	.03	.855

within groups comparisons with data collapsed over learning goal
Describers t(22) = 0.56, *p* = 0.583

test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	810.17	66	12.28		
VERB	156.25	1	156.25	12.73	.001
LRN_GOAL	102.26	2	51.13	4.17	.020
VERB BY LRN_GOAL	238.29	2	119.15	9.71	.000

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	343.33	66	5.20		
TRL BLOCK	16.00	1	16.00	3.08	.084
VERB BY TRL BLOCK	.00	1	.00	.00	1.000
LRN_GOAL BY TRL BLOCK	14.29	2	7.15	1.37	.260
VERB BY LRN_GOAL BY TRL BLOCK	35.37	2	17.69	3.40	.039

Between groups comparisons for total test phase score

'explain' pattern search vs 'explain' control task then,
vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	477.04167	331.58333	477.04167	15.07197	31.65092	.000
TOT_TST	408.37500	687.58333	408.37500	31.25379	13.06642	.002
TOT_TST	737.04167	414.91667	737.04167	18.85985	39.07994	.000
TOT_TST	352.66667	313.33333	352.66667	14.24242	24.76170	.000
TOT_TST	715.04167	291.58333	715.04167	13.25379	53.94999	.000

'explain' control task vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	2.66667	809.83333	2.66667	36.81061	.07244	.790
TOT_TST	28.16667	537.16667	28.16667	24.41667	1.15358	.294
TOT_TST	9.37500	435.58333	9.37500	19.79924	.47350	.499
TOT_TST	24.00000	413.83333	24.00000	18.81061	1.27588	.271

'explain' dual goal group vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	48.16667	893.16667	48.16667	40.59848	1.18642	.288
TOT_TST	2.04167	791.58333	2.04167	35.98106	.05674	.814
TOT_TST	42.66667	769.83333	42.66667	34.99242	1.21931	.281

'describe' pattern search vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	70.04167	518.91667	70.04167	23.58712	2.96949	.099
TOT_TST	.16667	497.16667	.16667	22.59848	.00738	.932

'describe' control task vs 'describe' dual goal group

TOT_TST	63.37500	395.58333	63.37500	17.98106	3.52454	.074

Within groups comparisons between first and last 15 trials

'explain' pattern search t(11) = -1.34, *p* = 0.207
'explain' control task t(11) = -2.82, *p* = 0.017
'explain' dual goal group t(11) = 1.18, *p* = 0.263
'describe' pattern search t(11) = -0.11, *p* = 0.915
'describe' control task t(11) = -0.46, *p* = 0.653
'describe' dual goal group t(11) = -1.28, *p* = 0.229

transfer between learning trials and test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	518.62	44	11.79		
VERB	12.76	1	12.76	1.08	.304
LRN GOAL	3.76	1	3.76	.32	.575
VERB BY LRN GOAL	15.84	1	15.84	1.34	.253

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	217.29	44	4.94		
TRL BLOCK	49.59	1	49.59	10.04	.003
VERB BY TRL BLOCK	1.26	1	1.26	.26	.616
LRN GOAL BY TRL BLOCK	2.34	1	2.34	.47	.494
VERB BY LRN GOAL BY TRL BLOCK	44.01	1	44.01	8.91	.005

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1784.79	44	40.56		
VERB	2.34	1	2.34	.06	.811
LRN GOAL	3.76	1	3.76	.09	.762
VERB BY LRN GOAL	3.01	1	3.01	.07	.787

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	723.63	44	16.45		
PHASE SCORE	455.01	1	455.01	27.67	.000
VERB BY PHASE SCORE	3.76	1	3.76	.23	.635
LRN GOAL BY PHASE SCORE	68.34	1	68.34	4.16	.048
VERB BY LRN GOAL BY PHASE SCORE	61.76	1	61.76	3.76	.059

Within groups comparisons between last half of learning phase and first half of test phase.

'explain' control task t(11) = -0.16, *p* = 0.879
'describe' dual goal group t(11) = -0.00, *p* = 1.000

Within groups comparisons between overall scores between learning and test phases

'describe' dual goal group t(11) = -0.45, *p* = 0.663

prediction questions analysis done on bottom chart of Figure 3.3

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	80460.31	64	1257.19		
VERB	27731.24	1	27731.24	22.06	.000
LRN_GOAL	14386.81	2	7193.40	5.72	.005
VERB BY LRN_GOAL	21370.76	2	10685.38	8.50	.001

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	83104.46	128	649.25		
QUE_TYPE	16710.91	2	8355.46	12.87	.000
VERB BY QUE_TYPE	78.87	2	39.44	.06	.941
LRN_GOAL BY QUE_TYPE	8146.01	4	2036.50	3.14	.017
VERB BY LRN_GOAL BY QUE_TYPE	2177.67	4	544.42	.84	.503

Between groups comparisons for total prediction questionnaire score

'explain' pattern search vs 'explain' control task then,
vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
PTOT_QUE	8492.69669	3422.09128	8492.69669	162.95673	52.11627	.000
PTOT_QUE	8822.91677	6627.02955	8822.91677	315.57284	27.95842	.000
PTOT_QUE	14169.8232	10773.3446	14169.8232	538.66723	26.30534	.000
PTOT_QUE	12946.6185	2992.53881	12946.6185	142.50185	90.85228	.000
PTOT_QUE	12249.5617	5342.95034	12249.5617	254.42621	48.14583	.000

'explain' control task vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
PTOT_QUE	3.29218	8412.75720	3.29218	382.39805	.00861	.927
PTOT_QUE	866.80005	12559.0722	866.80005	598.05106	1.44937	.242
PTOT_QUE	489.00463	4778.26646	489.00463	217.19393	2.25147	.148
PTOT_QUE	358.65484	7128.67798	358.65484	324.03082	1.10685	.304

'explain' dual goal group vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
PTOT_QUE	765.45812	15764.0105	765.45812	750.66717	1.01970	.324
PTOT_QUE	412.04990	7983.20473	412.04990	362.87294	1.13552	.298
PTOT_QUE	293.22274	10333.6163	293.22274	469.70983	.62426	.438

'describe' pattern search vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
PTOT_QUE	61.05998	12129.5197	61.05998	577.59618	.10571	.748
PTOT_QUE	119.23675	14479.9313	119.23675	689.52054	.17293	.682

'describe' control task vs 'describe' dual goal group

PTOT_QUE	10.08230	6699.12551	10.08230	304.50571	.03311	.857

Within group comparisons for question type with data collapsed over verbalisation

Pattern search subjects

New vs Old-correct	Z = -1.3628	2-Tailed P =	.1730
New vs Old-wrong	Z = -2.7923	2-Tailed P =	.0052
Old-correct vs Old-wrong	Z = -2.1004	2-Tailed P =	.0357

control task subjects

New vs Old-correct	Z = -2.9655	2-Tailed P =	.0030
New vs Old-wrong	Z = -2.7758	2-Tailed P =	.0055
Old-correct vs Old-wrong	Z = -1.0286	2-Tailed P =	.3037

dual goal subjects

New vs Old-correct	Z =	-3.0680	2-Tailed P =	.0022
New vs Old-wrong	Z =	-.0933	2-Tailed P =	.9256
Old-correct vs Old-wrong	Z =	-2.6693	2-Tailed P =	.0076

'explain' patterns search subjects

New vs Old-correct	Z =	-.1048	2-Tailed P =	.9165
New vs Old-wrong	Z =	-1.5724	2-Tailed P =	.1159
Old-correct vs Old-wrong	Z =	-1.3416	2-Tailed P =	.1797

'describe' pattern search subjects

New vs Old-correct	Z =	-1.6058	2-Tailed P =	.1083
New vs Old-wrong	Z =	-2.2424	2-Tailed P =	.0249
Old-correct vs Old-wrong	Z =	-1.5724	2-Tailed P =	.1159

<u>Prediction questions - Analysis done when prediction questions defined by the last two elements rather than all three.</u>

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	86523.98	62	1395.55		
VERB	19026.57	1	19026.57	13.63	.000
LRN_GOAL	14110.67	2	7055.34	5.06	.009
VERB BY LRN_GOAL	19631.14	2	9815.57	7.03	.002

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	98584.67	124	795.04		
QUE_TYPE	14482.31	2	7241.16	9.11	.000
VERB BY QUE_TYPE	149.05	2	74.53	.09	.911
LRN_GOAL BY QUE_TYPE	7893.23	4	1973.31	2.48	.047
VERB BY LRN_GOAL BY QUE_TYPE	1803.87	4	450.97	.57	.687

<u>Between groups comparisons for total prediction questionnaire score</u>

'explain' pattern search vs 'explain' control task then,
vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P2TOT_QU	7746.39918	3934.92798	7746.39918	196.74640	39.37251	.000
P2TOT_QU	8690.63986	7353.12851	8690.63986	387.00676	22.45604	.000
P2TOT_QU	11893.9695	11634.1611	11893.9695	612.32427	19.42430	.000
P2TOT_QU	10819.3509	3308.38477	10819.3509	165.41924	65.40564	.000
P2TOT_QU	10707.1759	7436.03395	10707.1759	371.80170	28.79808	.000

'explain' control task vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P2TOT_QU	53.28811	8875.40217	53.28811	422.63820	.12608	.726
P2TOT_QU	570.05482	13156.4347	570.05482	626.49689	.90991	.351
P2TOT_QU	281.68724	4830.65844	281.68724	219.57538	1.28287	.270
P2TOT_QU	262.97582	8958.30761	262.97582	407.19580	.64582	.430

'explain' dual goal group vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P2TOT_QU	263.31369	16574.6352	263.31369	828.73176	.31773	.579
P2TOT_QU	83.07877	8248.85896	83.07877	392.80281	.21150	.650
P2TOT_QU	73.27716	12376.5081	73.27716	589.35753	.12433	.728

'describe' pattern search vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P2TOT_QU	55.66942	12529.8915	55.66942	596.66150	.09330	.763
P2TOT_QU	64.25212	16657.5407	64.25212	793.21622	.08100	.779

'describe' control task vs 'describe' dual goal group

P2TOT_QU	.32150	8331.76440	.32150	378.71656	.00085	.977

<u>Within group comparisons for question type with data collapsed over verbalisation</u>

Pattern search subjects		
New vs Old-correct	Z = -1.2159	2-Tailed P = .2240
New vs Old-wrong	Z = -2.7147	2-Tailed P = .0066
Old-correct vs Old-wrong	Z = -1.5401	2-Tailed P = .1235
control task subjects		
New vs Old-correct	Z = -2.8849	2-Tailed P = .0039
New vs Old-wrong	Z = -2.3000	2-Tailed P = .0214
Old-correct vs Old-wrong	Z = -.8714	2-Tailed P = .3835
dual goal subjects		
New vs Old-correct	Z = -2.3520	2-Tailed P = .0187
New vs Old-wrong	Z = -.0373	2-Tailed P = .9702
Old-correct vs Old-wrong	Z = -2.2026	2-Tailed P = .0276
'explain' patterns search subjects		
New vs Old-correct	Z = -.1048	2-Tailed P = .9165
New vs Old-wrong	Z = -1.5724	2-Tailed P = .1159
Old-correct vs Old-wrong	Z = -1.3416	2-Tailed P = .1797
'describe' pattern search subjects		
New vs Old-correct	Z = -1.7821	2-Tailed P = .0747
New vs Old-wrong	Z = -2.2424	2-Tailed P = .0249
Old-correct vs Old-wrong	Z = -.9297	2-Tailed P = .3525

<u>Spearman rank correlations between total predictions question scores and number of correct test and learning trials.</u>

	'exp' PS	'exp' CT	'exp' DG	'des' PS	'des' CT	'des' DG
TOT_TST	.5762	-.0389	.6941	.6028	-.2355	.4000
	N(10)	N(12)	N(11)	N(11)	N(12)	N(12)
	Sig .081	Sig .904	Sig .018	Sig .050	Sig .461	Sig .198
TOT_LRN		.1943	.4470		-.6431	.2250
		N(12)	N(11)		N(12)	N(12)
		Sig .545	Sig .168		Sig .024	Sig .482
	P2TOT_QU	P2TOT_QU	P2TOT_QU	P2TOT_QU	P2TOT_QU	P2TOT_QU

EXPERIMENT 3'S F TABLES AND ADDITIONAL STATISTICS

Prediction questions - Analysis done when prediction questions have Old-wrong and Old-correct questions refined so that there are no Old-wrong situations that were Old-correct situations in the learning phase and also so that there were no Old-correct situations that were Old-wrong situations in the learning phase.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	79381.52	64	1240.34		
VERB	25036.44	1	25036.44	20.19	.000
LRN_GOAL	12341.80	2	6170.90	4.98	.010
VERB BY LRN_GOAL	22019.54	2	11009.77	8.88	.000

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	80418.97	128	628.27		
QUE_TYPE	19413.10	2	9706.55	15.45	.000
VERB BY QUE_TYPE	72.24	2	36.12	.06	.944
LRN_GOAL BY QUE_TYPE	8130.37	4	2032.59	3.24	.014
VERB BY LRN_GOAL BY QUE_TYPE	1918.76	4	479.69	.76	.551

Between groups comparisons for total prediction questionnaire score

'explain' pattern search vs 'explain' control task then,
vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P3TOT_QU	8492.69669	3422.09128	8492.69669	162.95673	52.11627	.000
P3TOT_QU	8574.66127	6685.67153	8574.66127	318.36531	26.93340	.000
P3TOT_QU	14453.2127	10860.3255	14453.2127	543.01627	26.61654	.000
P3TOT_QU	11739.4269	3422.19416	11739.4269	162.96163	72.03798	.000
P3TOT_QU	12249.5617	5342.95034	12249.5617	254.42621	48.14583	.000

'explain' control task vs 'explain' dual goal group then,
vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P3TOT_QU	.20576	8471.39918	.20576	385.06360	.00053	.982
P3TOT_QU	939.50780	12646.0531	939.50780	602.19301	1.56014	.225
P3TOT_QU	274.12551	5207.92181	274.12551	236.72372	1.15800	.294
P3TOT_QU	358.65484	7128.67798	358.65484	324.03082	1.10685	.304

'explain' dual goal group vs 'describe' pattern search then,
vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P3TOT_QU	912.50834	15909.6334	912.50834	757.60159	1.20447	.285
P3TOT_QU	259.31070	8471.50206	259.31070	385.06828	.67341	.421
P3TOT_QU	341.67953	10392.2582	341.67953	472.37537	.72332	.404

'describe' pattern search vs 'describe' control task then,
vs 'describe' dual goal group

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
P3TOT_QU	209.05065	12646.1560	209.05065	602.19790	.34715	.562
P3TOT_QU	147.12432	14566.9122	147.12432	693.66248	.21210	.650

'describe' control task vs 'describe' dual goal group

P3TOT_QU	5.67130	7128.78086	5.67130	324.03549	.01750	.896

Within group comparisons for question type with data collapsed over verbalisation

Pattern search subjects

New vs Old-correct	Z =	-1.2579	2-Tailed P =	.2084
New vs Old-wrong	Z =	-2.7923	2-Tailed P =	.0052
Old-correct vs Old-wrong	Z =	-2.2404	2-Tailed P =	.0251

EXPERIMENT 3'S F TABLES AND ADDITIONAL STATISTICS

control task subjects

New vs Old-correct	Z =	-3.0871	2-Tailed P =	.0020
New vs Old-wrong	Z =	-2.8894	2-Tailed P =	.0039
Old-correct vs Old-wrong	Z =	-1.2012	2-Tailed P =	.2297

dual goal subjects

New vs Old-correct	Z =	-3.0680	2-Tailed P =	.0022
New vs Old-wrong	Z =	-.0348	2-Tailed P =	.9723
Old-correct vs Old-wrong	Z =	-2.6320	2-Tailed P =	.0085

'explain' patterns search subjects

New vs Old-correct	Z =	-.1048	2-Tailed P =	.9165
New vs Old-wrong	Z =	-1.5724	2-Tailed P =	.1159
Old-correct vs Old-wrong	Z =	-1.3416	2-Tailed P =	.1797

'describe' pattern search subjects

New vs Old-correct	Z =	-1.4368	2-Tailed P =	.1508
New vs Old-wrong	Z =	-2.2424	2-Tailed P =	.0249
Old-correct vs Old-wrong	Z =	-1.7821	2-Tailed P =	.0747

Spearman rank correlations between total predictions question scores and number of correct test and learning trials.

	'exp' PS	'exp' CT	'exp' DG	'des' PS	'des' CT	'des' DG
TOT_TST	.8205	.4621	.7088	.6005	-.1144	.4298
	N(11)	N(12)	N(12)	N(11)	N(12)	N(12)
	Sig .002	Sig .130	Sig .010	Sig .051	Sig .723	Sig .163
TOT_LRN	.	.1162	.4832		-.5469	.2847
	N(11)	N(12)	N(12)		N(12)	N(12)
	Sig .	Sig .719	Sig .112		Sig .066	Sig .370
	P3TOT_QU	P3TOT_QU	P3TOT_QU	P3TOT_QU	P3TOT_QU	P3TOT_QU

Verbalising vs not verbalising: Comparisons with the data from experiment1

Comparisons done for total learning and test phases and for total prediction question scores and for the Old-wrong, Old-correct and new prediction question scores.

Experiment 1's Pattern search subjects vs 'describe' pattern search subjects.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	341.74556	878.14015	341.74556	26.61031	12.84260	.001
TOT_QUE1	6409.64783	32247.1746	6409.64783	977.18711	6.55928	.015
OW1	2103.18212	37125.3893	2103.18212	1125.01180	1.86948	.181
OC1	4666.78692	42084.0067	4666.78692	1275.27293	3.65944	.064
NEW1	13238.4848	45001.5152	13238.4848	1363.68228	9.70790	.004

Experiment 1's Pattern search subjects vs 'explain' pattern search subjects.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	168.29102	705.59470	168.29102	21.38166	7.87081	.008
TOT_QUE1	3521.43287	23110.1937	3521.43287	700.30890	5.02840	.032
OW1	5785.76509	28844.0762	5785.76509	874.06291	6.61939	.015
OC1	3385.86340	33033.5017	3385.86340	1001.01520	3.38243	.075
NEW1	2994.32900	32274.2424	2994.32900	978.00735	3.06166	.089

Experiment 1's Control task subjects vs 'describe' control task subjects.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_LRN	196.68056	560.54167	196.68056	16.48652	11.92978	.001
TOT_TST	56.88889	756.66667	56.88889	22.25490	2.55624	.119
TOT_QUE1	261.59674	8153.81125	261.59674	239.81798	1.09081	.304
OW1	22.22222	22102.0833	22.22222	650.06127	.03418	.854
OC1	1653.12500	19976.7361	1653.12500	587.55106	2.81359	.103
NEW1	5.55556	21183.3333	5.55556	623.03922	.00892	.925

Experiment 1's Control task subjects vs 'explain' control task subjects.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_LRN	115.01389	608.54167	115.01389	17.89828	6.42597	.016
TOT_TST	16.05556	774.91667	16.05556	22.79167	.70445	.407
TOT_QUE1	1739.58762	8583.36372	1739.58762	252.45187	6.89077	.013
OW1	2334.72222	20559.7222	2334.72222	604.69771	3.86097	.058
OC1	3542.01389	18421.1806	3542.01389	541.79943	6.53750	.015
NEW1	138.88889	28083.3333	138.88889	825.98039	.16815	.684

Experiment 1's dual goal subjects vs 'describe' control task subjects.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_LRN	100.34722	668.20833	100.34722	19.65319	5.10590	.030
TOT_TST	30.68056	484.87500	30.68056	14.26103	2.15136	.152
TOT_QUE1	1219.70244	10933.7720	1219.70244	321.58153	3.79282	.060
OW1	555.55556	24385.4167	555.55556	717.21814	.77460	.385
OC1	1918.55710	23513.7731	1918.55710	691.58156	2.77416	.105
NEW1	1422.22222	20533.3333	1422.22222	603.92157	2.35498	.134

Experiment 1's dual goal subjects vs 'explain' control task subjects.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_LRN	39.01389	664.87500	39.01389	19.55515	1.99507	.167
TOT_TST	171.12500	880.87500	171.12500	25.90809	6.60508	.015
TOT_QUE1	2991.76611	12217.8512	2991.76611	359.34857	8.32553	.007
OW1	3068.05556	31829.1667	3068.05556	936.15196	3.27731	.079
OC1	5659.29784	28082.2917	5659.29784	825.94975	6.85187	.013
NEW1	1141.35802	26787.9630	1141.35802	787.88126	1.44864	.237

APPENDIX 4

EXPERIMENT 4'S F TABLES AND ADDITIONAL STATISTICS

learning trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	261.12	44	5.93		
LRN GOAL	.51	1	.51	.09	.771
SEC_TASK	14.26	1	14.26	2.40	.128
LRN GOAL BY SEC_TASK	11.34	1	11.34	1.91	.174

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	302.46	44	6.87		
TRL BLOCK	94.01	1	94.01	13.68	.001
LRN GOAL BY TRL BLOCK	.01	1	.01	.00	.969
SEC_TASK BY TRL BLOCK	.26	1	.26	.04	.847
LRN GOAL BY SEC_TASK BY TRL BLOCK	8.76	1	8.76	1.27	.265

test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	331.00	66	5.02		
SEC_TASK	.69	1	.69	.14	.711
LRN GOAL	14.29	2	7.15	1.42	.248
SEC_TASK BY LRN GOAL	4.76	2	2.38	.47	.624

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	660.17	66	10.00		
TRL BLOCK	36.00	1	36.00	3.60	.062
SEC_TASK BY TRL BLOCK	7.11	1	7.11	.71	.402
LRN GOAL BY TRL BLOCK	92.63	2	46.31	4.63	.013
SEC_TASK BY LRN GOAL BY TRL BLOCK	7.10	2	3.55	.35	.703

Within groups comparisons between first and last half of test phase, with data collapsed over secondary task.

Pattern search subjects	$t(22) = -4.48$, $p = 0.000$
Control task subjects	$t(22) = -0.13$, $p = 0.895$
dual goal subjects	$t(22) = -0.37$, $p = 0.717$

transfer between learning and test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	455.46	44	10.35		
SEC_TASK	10.01	1	10.01	.97	.331
LRN GOAL	1.76	1	1.76	.17	.682
SEC_TASK BY LRN GOAL	12.76	1	12.76	1.23	.273

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	264.13	44	6.00		
TRL BLOCK	1.76	1	1.76	.29	.591
SEC_TASK BY TRL BLOCK	1.26	1	1.26	.21	.649
LRN GOAL BY TRL BLOCK	3.76	1	3.76	.63	.433
SEC_TASK BY LRN GOAL BY TRL BLOCK	7.59	1	7.59	1.27	.267

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	548.46	44	12.46		
SEC_TASK	5.51	1	5.51	.44	.510
LRN GOAL	.01	1	.01	.00	.977
SEC_TASK BY LRN GOAL	.01	1	.01	.00	.977

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	370.79	44	8.43		
PHASE SCORE	137.76	1	137.76	16.35	.000
SEC_TASK BY PHASE SCORE	14.26	1	14.26	1.69	.200
LRN GOAL BY PHASE SCORE	7.59	1	7.59	.90	.348
SEC_TASK BY LRN GOAL BY PHASE SCORE	27.09	1	27.09	3.22	.080

Analysis of prediction questions in Figure 4.3

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	49843.90	66	755.21		
SEC_TASK	594.46	1	594.46	.79	.378
LRN GOAL	2075.03	2	1037.51	1.37	.260
SEC_TASK BY LRN GOAL	1093.70	2	546.85	.72	.489

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	82051.70	132	621.60		
QUE_TYPE	22629.35	2	11314.67	18.20	.000
SEC_TASK BY QUE_TYPE	352.34	2	176.17	.28	.754
LRN GOAL BY QUE_TYPE	3745.19	4	936.30	1.51	.204
SEC_TASK BY LRN GOAL BY QUE_TYPE	2717.72	4	679.43	1.09	.363

Spearman Rank correlation coefficients for all groups

V1= verbalising only in learning phase, V2 = verbalising only in test phase.

	V1 PS	V1 CT	V1 DG	V2 PS	V2 CT	V2 DG
TOT_LRN		.1875 N(12) Sig .560	-.1358 N(12) Sig .674		.1135 N(12) Sig .725	.4655 N(12) Sig .127
TOT_TST	.0142 N(12) Sig .965 PTOT_QUE	-.2522 N(12) Sig .429 PTOT_QUE	.4425 N(12) Sig .150 PTOT_QUE	-.1213 N(12) Sig .707 PTOT_QUE	.2968 N(12) Sig .349 PTOT_QUE	.1416 N(12) Sig .661 PTOT_QUE

Prediction questions - Analysis done when prediction questions defined by the last two elements rather than all three.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	58302.39	66	883.37		
SEC_TASK	350.12	1	350.12	.40	.531
LRN_GOAL	2533.41	2	1266.71	1.43	.246
SEC_TASK BY LRN_GOAL	2206.56	2	1103.28	1.25	.293

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	100784.88	132	763.52		
QUE_TYPE	21905.17	2	10952.58	14.34	.000
SEC_TASK BY QUE_TYPE	46.53	2	23.26	.03	.970
LRN_GOAL BY QUE_TYPE	3188.27	4	797.07	1.04	.387
SEC_TASK BY LRN_GOAL BY QUE_TYPE	2910.34	4	727.58	.95	.436

Comparison for Old-wrong vs Old-correct, Old-correct vs new, Old-wrong vs new, with data collapsed over secondary task and learning goal

Old-correct vs Old-wrong	Z =	-3.3821	2-Tailed P =	.0007
Old-correct vs new	Z =	-4.2289	2-Tailed P =	.0000
Old-wrong vs new	Z =	-.9992	2-Tailed P =	.3177

Spearman Rank correlation coefficients for all groups
V1= verbalising only in learning phase, V2 = verbalising only in test phase.

	V1 PS	V1 CT	V1 DG	V2 PS	V2 CT	V2 DG
TOT_LRN		.4171	-.0159		.2642	.3310
		N(12)	N(12)		N(12)	N(12)
		Sig .177	Sig .961		Sig .407	Sig .293
TOT_TST	-.1150	-.1071	.5806	-.0976	.1572	.0283
	N(12)	N(12)	N(12)	N(12)	N(12)	N(12)
	Sig .722	Sig .740	Sig .048	Sig .763	Sig .626	Sig .930
	TOT_QU2	TOT_QU2	TOT_QU2	TOT_QU2	TOT_QU2	TOT_QU2

Prediction questions - Analysis done when prediction questions have Old-wrong and Old-correct questions refined so that there are no Old-wrong situations that were Old-correct situations in the learning phase and also so that there were no Old-correct situations that were Old-wrong situations in the learning phase.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	52467.36	66	794.96		
SEC_TASK	876.04	1	876.04	1.10	.298
LRN_GOAL	1990.53	2	995.27	1.25	.293
SEC_TASK BY LRN_GOAL	1496.60	2	748.30	.94	.395

- -

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	91289.35	132	691.59		
QUE_TYPE	27581.04	2	13790.52	19.94	.000
SEC_TASK BY QUE_TYPE	540.82	2	270.41	.39	.677
LRN_GOAL BY QUE_TYPE	4221.81	4	1055.45	1.53	.198
SEC_TASK BY LRN_GOAL BY QUE_TYPE	3394.75	4	848.69	1.23	.302

Comparison for Old-wrong vs Old-correct, Old-correct vs new, Old-wrong vs new, with data collapsed over secondary task and learning goal

Old-correct vs Old-wrong	Z =	-3.9885	2-Tailed P =	.0001
Old-correct vs new	Z =	-4.6109	2-Tailed P =	.0000
Old-wrong vs new	Z =	-.7289	2-Tailed P =	.4660

Spearman Rank correlation coefficients for all groups
V1= verbalising only in learning phase, V2 = verbalising only in test phase.

	V1 PS	V1 CT	V1 DG	V2 PS	V2 CT	V2 DG
TOT_LRN		.2527	-.1093		.1818	.3598
		N(12)	N(12)		N(12)	N(12)
		Sig .428	Sig .735		Sig .572	Sig .251
TOT_TST	.2368	-.3599	.5965	-.1649	.2096	.0705
	N(12)	N(12)	N(12)	N(12)	N(12)	N(12)
	Sig .459	Sig .251	Sig .041	Sig .609	Sig .513	Sig .828
	P3TOT_QU	P3TOT_QU	P3TOT_QU	P3TOT_QU	P3TOT_QU	P3TOT_QU

Random number generation

For the learning phase, with data collapsed over secondary task

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
LP_RNG	.00304	.18865	.00152	.00273	.55596	.576
N_RNG_LP	941.36111	257149.083	470.68056	3726.79831	.12630	.882

for the test phase,

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TP_RNG	.00429	.06434	.00214	.00195	1.09894	.345
N_RNG_TP	7432.16667	86266.8333	3716.08333	2614.14646	1.42153	.256

Spearman Rank correlation coefficients of LEARNING PHASE RNG (LP RNG) and Number of digits produced during LEARNING PHASE (N RNG LP)with total learning and test phase scores and total prediction question score. (ptot que = total prediction questionnaire score as from Figure 4.3, tot qu2 = total prediction questionnaire score as from the second analysis, above, on the prediction question scores, p3tot = total prediction questionnaire score as from the third analysis, above, on the prediction question scores,

	TOT_LRN	TOT_TST	PTOT_QUE	TOT_QU2	P3TOT_QU
LP_RNG	.0844	.0791	.0739	.1644	.0918
	N(72)	N(72)	N(72)	N(72)	N(72)
	Sig .481	Sig .509	Sig .537	Sig .168	Sig .443
N_RNG_LP	.0368	.1210	.1244	.1792	.1087
	N(72)	N(72)	N(72)	N(72)	N(72)
	Sig .759	Sig .311	Sig .298	Sig .132	Sig .363

Spearman Rank correlation coefficients of TEST PHASE RNG (TP RNG) and Number of digits produced during TEST PHASE (N RNG TP)with total learning and test phase scores and total prediction question score. (ptot que = total prediction questionnaire score as from Figure 4.3, tot qu2 = total prediction questionnaire score as from the second analysis, above, on the prediction question scores, p3tot = total prediction questionnaire score as from the third analysis, above, on the prediction question scores,

	TOT_LRN	TOT_TST	PTOT_QUE	TOT_QU2	P3TOT_QU
TP_RNG	-.1155	.0913	-.0878	-.0402	-.0946
	N(36)	N(36)	N(36)	N(36)	N(36)
	Sig .502	Sig .597	Sig .611	Sig .816	Sig .583
N_RNG_TP	-.0407	.0073	-.0473	.0314	-.0318
	N(36)	N(36)	N(36)	N(36)	N(36)
	Sig .814	Sig .966	Sig .784	Sig .856	Sig .854

Comparisons with the dual goal group from Experiment 1, data is collapsed across learning goal and secondary task. For the comparison with the learning phase data the pattern search subjects are omitted.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	5.55556	978.94444	5.55556	10.41430	.53345	.467
PTOT_QUE	10.62124	24531.0666	10.62124	260.96879	.04070	.841
PNEW	22.22222	51111.1111	22.22222	543.73522	.04087	.840
POW	.34722	58473.6111	.34722	622.05969	.00056	.981
POC	47.26080	76873.9198	47.26080	817.80766	.05779	.811
TOT_LRN	11.11111	801.87500	11.11111	11.45536	.96995	.328

APPENDIX 5

EXPERIMENT 5'S F TABLES AND ADDITIONAL STATISTICS

learning trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	197.96	22	9.00		
LRN GOAL	7.52	1	7.52	.84	.371

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	112.46	22	5.11		
TRL BLOCK	11.02	1	11.02	2.16	.156
LRN GOAL BY TRL BLOCK	13.02	1	13.02	2.55	.125

test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	652.79	33	19.78		
LRN GOAL	250.58	2	125.29	6.33	.005

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	101.79	33	3.08		
TRL BLOCK	.68	1	.68	.22	.642
LRN GOAL BY TRL BLOCK	.03	2	.01	.00	.996

Between groups comparisons for the total test phase scores

Pattern search subjects vs control task subjects, then
vs dual goal subjects, then
control task subjects vs dual goal subjects

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	495.04167	804.58333	495.04167	36.57197	13.53610	.001
TOT_TST	176.04167	935.91667	176.04167	42.54167	4.13810	.054
TOT_TST	80.66667	870.66667	80.66667	39.57576	2.03828	.167

Transfer between learning and test phase transfer

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	241.25	22	10.97		
LRN GOAL	.00	1	.00	.00	1.000

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	211.58	22	9.62		
TRL BLOCK	14.08	1	14.08	1.46	.239
LRN GOAL BY TRL BLOCK	40.33	1	40.33	4.19	.053

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	638.79	22	29.04		
LRN GOAL	13.02	1	13.02	.45	.510

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	627.79	22	28.54		
PHASE SCORE	105.02	1	105.02	3.68	.068
LRN GOAL BY PHASE SCORE	82.69	1	82.69	2.90	.103

EXPERIMENT 5'S F TABLES AND ADDITIONAL STATISTICS

recognition of instances, analysis on data in figure 5.3

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	16476.77	33	499.30		
LRN_GOAL	993.06	2	496.53	.99	.381

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	8418.18	33	255.10		
INSTANCE TYPE	6368.64	1	6368.64	24.97	.000
LRN_GOAL BY INSTANCE TYPE	301.91	2	150.96	.59	.559

Within group comparisons

Pattern search subjects
Old-correct vs Old-wrong $t(11) = -0.91$, $p = 0.38$
Old-correct vs New-legal $t(11) = -6.24$, $p = 0.000$
Old-wrong vs New-legal $t(11) = -5.16$, $p = 0.000$

control task subjects
Old-correct vs Old-wrong $t(11) = -0.72$, $p = 0.489$

dual goal subjects
Old-correct vs Old-wrong $t(11) = -2.73$, $p = 0.02$

Comparisons to chance (a value of 50%)

Pattern search subjects
New-illegal $t(11) = 8.68$, $p = 0.000$
New-legal $t(11) = 1.91$, $p = 0.082$
Old-wrong $t(11) = 5.02$, $p = 0.000$
Old-correct $t(11) = 9.29$, $p = 0.000$

control task subjects
New-illegal $t(11) = 4.97$, $p = 0.000$
New-legal $t(11) = 2.19$, $p = 0.051$
Old-wrong $t(11) = 3.89$, $p = 0.003$
Old-correct $t(11) = 3.35$, $p = 0.006$

dual goal subjects
New-illegal $t(11) = 6.91$, $p = 0.000$
New-legal $t(11) = 2.16$, $p = 0.054$
Old-wrong $t(11) = 5.90$, $p = 0.000$
Old-correct $t(11) = 5.44$, $p = 0.000$

Recognition of instances with Old-wrong refined so that none of the Old-wrong instances were actually Old-correct instances in the learning phase.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	17508.29	33	530.55		
LRN_GOAL	1011.15	2	505.57	.95	.396

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	9347.19	33	283.25		
INSTANCE TYPE	6430.66	1	6430.66	22.70	.000
LRN_GOAL BY INSTANCE TYPE	310.85	2	155.43	.55	.583

comparisons of instance types including refined Old-wrong instances

New-illegal vs Old-wrong $t(35)= -13.07$, $p = 000$
New-legal vs Old-wrong $t(35)= -7.04$, $p = 000$
Old-correct vs Old-wrong $t(35)= 2.13$, $p = 040$

Within group comparisons

Pattern search subjects
Old-correct vs Old-wrong $t(11) = -0.90$, $p = 0.385$
Old-wrong vs New-legal $t(11) = -5.17$, $p = 0.000$

control task subjects
Old-correct vs Old-wrong $t(11) = -0.73$, $p = 0.483$

dual goal subjects
Old-correct vs Old-wrong $t(11) = -2.28$, $p = 0.044$

Comparisons to chance (a value of 50%)

Pattern search subjects
Old-wrong $t(11) = -5.04$, $p = 0.000$

control task subjects
Old-wrong $t(11) = 3.87$, $p = 0.003$

dual goal subjects
Old-wrong $t(11) = -3.24$, $p = 0.008$

EXPERIMENT 6A & 6B'S F TABLES AND ADDITIONAL STATISTICS

EXPERIMENT 6a

learning trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	129.92	22	5.91		
TASK	21.33	1	21.33	3.61	.071

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	285.58	22	12.98		
TRL BLOCK	80.08	1	80.08	6.17	.021
TASK BY TRL BLOCK	16.33	1	16.33	1.26	.274

test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	530.38	44	12.05		
TASK	.51	1	.51	.04	.838
LRN GOAL	86.26	1	86.26	7.16	.010
TASK BY LRN GOAL	.84	1	.84	.07	.793

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	281.88	44	6.41		
TRL BLOCK	173.34	1	173.34	27.06	.000
TASK BY TRL BLOCK	21.09	1	21.09	3.29	.076
LRN GOAL BY TRL BLOCK	4.59	1	4.59	.72	.402
TASK BY LRN GOAL BY TRL BLOCK	7.59	1	7.59	1.19	.282

transfer between learning and test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	142.33	22	6.47		
TASK	40.33	1	40.33	6.23	.021

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	197.33	22	8.97		
TRL BLOCK	56.33	1	56.33	6.28	.020
TASK BY TRL BLOCK	5.33	1	5.33	.59	.449

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	153.46	22	6.98		
TASK	6.02	1	6.02	.86	.363

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	157.29	22	7.15		
PHASE SCORE	.52	1	.52	.07	.790
TASK BY PHASE SCORE	22.69	1	22.69	3.17	.089

prediction questions, analysis done on data in Figure 6.3

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	59976.93	44	1363.11		
TASK	1.56	1	1.56	.00	.973
LRN GOAL	56208.51	1	56208.51	41.24	.000
TASK BY LRN GOAL	1747.70	1	1747.70	1.28	.264

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	33696.91	88	382.92		
QUE_TYPE	3710.92	2	1855.46	4.85	.010
TASK BY QUE_TYPE	1404.98	2	702.49	1.83	.166
LRN GOAL BY QUE_TYPE	5679.98	2	2839.99	7.42	.001
TASK BY LRN GOAL BY QUE_TYPE	844.25	2	422.13	1.10	.337

Within group comparisons

control task subjects

Old-wrong vs Old-correct	Z = -1.3208	2-Tailed P =	.1866
Old-wrong vs new	Z = -3.6213	2-Tailed P =	.0003
Old-correct vs new	Z = -2.9377	2-Tailed P =	.0033

pattern search subjects

Old-wrong vs Old-correct	Z = -.2097	2-Tailed P =	.8339
Old-wrong vs new	Z = -.4137	2-Tailed P =	.6791
Old-correct vs new	Z = -.0871	2-Tailed P =	.9306

Prediction questions, analysis done on data where situations defined by last two elements only.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	63429.31	43	1475.10		
TASK_ORD	133.37	1	133.37	.09	.765
LRN_GOAL	50230.45	1	50230.45	34.05	.000
TASK_ORD BY LRN_GOAL	2959.14	1	2959.14	2.01	.164

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	38899.41	86	452.32		
QUE_TYPE	3134.53	2	1567.26	3.46	.036
TASK_ORD BY QUE_TYPE	1590.30	2	795.15	1.76	.179
LRN_GOAL BY QUE_TYPE	4394.23	2	2197.11	4.86	.010
TASK_ORD BY LRN_GOAL BY QUE_TYPE	1756.22	2	878.11	1.94	.150

Within group comparisons

control task subjects

Old-wrong vs Old-correct	Z = -1.3798	2-Tailed P =	.1677
Old-wrong vs new	Z = -3.1629	2-Tailed P =	.0016
Old-correct vs new	Z = -2.3542	2-Tailed P =	.0186

pattern search subjects

Old-wrong vs Old-correct	Z = -.1704	2-Tailed P =	.8647
Old-wrong vs new	Z = -.1810	2-Tailed P =	.8564
Old-correct vs new	Z = -.1894	2-Tailed P =	.8498

Spearman rank correlation coefficients between total learning, test, and transfer phase scores and total prediction questions.

Pi = person interaction task, F= factory task , psc = pattern search, ctk= control task

	pi-psc	f-psc	pi-ctk	f-ctk
TOT_LRN			-.1522	.2018
			N(12)	N(12)
			Sig .637	Sig .529
TOT_TST	.8283	.7983	-.2421	.2380
	N(12)	N(11)	N(12)	N(12)
	Sig .001	Sig .003	Sig .448	Sig .456
TOT_TRA	.6121	.8246	.1637	.4414
	N(12)	N(11)	N(12)	N(12)
	Sig .034	Sig .002	Sig .611	Sig .151
	TOT_QUE2	TOT_QUE2	TOT_QUE2	TOT_QUE2

prediction questions done on data where overlap between learning and test phase is removed - so Old-wrong is not actually learning phase Old-correct and vice versa

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	60083.18	44	1365.53		
TASK_ORD	11.11	1	11.11	.01	.929
LRN_GOAL	55355.63	1	55355.63	40.54	.000
TASK_ORD BY LRN_GOAL	1448.23	1	1448.23	1.06	.309

- -

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	37396.91	88	424.96		
QUE_TYPE	3476.50	2	1738.25	4.09	.020
TASK_ORD BY QUE_TYPE	1445.95	2	722.97	1.70	.188
LRN_GOAL BY QUE_TYPE	5309.76	2	2654.88	6.25	.003
TASK_ORD BY LRN_GOAL BY QUE_TYPE	700.50	2	350.25	.82	.442

Within group comparisons

control task subjects

Old-wrong vs Old-correct	Z = -.8766	2-Tailed P =	.3807
Old-wrong vs new	Z = -3.4719	2-Tailed P =	.0005
Old-correct vs new	Z = -2.9377	2-Tailed P =	.0033

pattern search subjects

Old-wrong vs Old-correct	Z = .0000	2-Tailed P =	1.0000
Old-wrong vs new	Z = -.4137	2-Tailed P =	.6791
Old-correct vs new	Z = -.2831	2-Tailed P =	.7771

Spearman rank correlation coefficients between total learning, test, and transfer phase scores and total prediction questions.

Pi = person interaction task, F= factory task , psc = pattern search, ctk= control task

	pi-psc	f-psc	pi-ctk	f-ctk
TOT_LRN	.		.0265	.2159
	N(12)		N(12)	N(12)
	Sig .		Sig .935	Sig .500
TOT_TST	.7965	.8811	-.3286	.2629
	N(12)	N(12)	N(12)	N(12)
	Sig .002	Sig .000	Sig .297	Sig .409
TOT_TRA	.5908	.7496	.2918	.4307
	N(12)	N(12)	N(12)	N(12)
	Sig .043	Sig .005	Sig .357	Sig .162
	TOT_QUE3	TOT_QUE3	TOT_QUE3	TOT_QUE3

transfer between tasks

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1631.92	44	37.09		
TASK_ORD	60.17	1	60.17	1.62	.209
LRN GOAL	541.50	1	541.50	14.60	.000
TASK_ORD BY LRN GOAL	26.04	1	26.04	.70	.407

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	396.58	44	9.01		
PHASE SCORE	88.17	1	88.17	9.78	.003
TASK_ORD BY PHASE SCORE	40.04	1	40.04	4.44	.041
LRN GOAL BY PHASE SCORE	22.04	1	22.04	2.45	.125
TASK_ORD BY LRN GOAL BY PHASE SCORE	48.17	1	48.17	5.34	.026

within groups comparisons comparing overall performance between transfer and test phase

For the pattern search group tran. to the factory task $t(11) = 2.48, p = 0.031$
For the pattern search group tran. to the person task $t(11) = 1.21, p = 0.250$
For the control task group tran. to the factory task $t(11) = 2.83, p = 0.016$
For the control task group tran. to the person task $t(11) = 3.35, p = 0.007$

Between groups comparisons for total transfer phase score

pit-f = person interaction task to factory task, f-pit = factory task to person interaction task. Cts= control task subjects, psc=pattern search subjects

pit-f cts vs pit-f psc, then,
vs f-pit cts, then
vs f-pit psc,

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TRA	63.37500	234.58333	63.37500	10.66288	5.94352	.023
TOT_TRA	1.04167	98.58333	1.04167	4.48106	.23246	.634
TOT_TRA	442.04167	745.91667	442.04167	33.90530	13.03754	.002

pit-f psc vs f-pit cts then
vs f-pit psc,

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TRA	48.16667	221.83333	48.16667	10.08333	4.77686	.040
TOT_TRA	170.66667	869.16667	170.66667	39.50758	4.31985	.050

f-pit cts vs f-pit psc,

TOT_TRA	400.16667	733.16667	400.16667	33.32576	12.00773	.002

similarity questions

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	146.28	44	3.32		
TASK_ORD	6.25	1	6.25	1.88	.177
LRN GOAL	42.25	1	42.25	12.71	.001
TASK_ORD BY LRN GOAL	.11	1	.11	.03	.856

Tests involving 'SIM_QUES' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	36.06	88	.41		
SIM_QUES	14.26	2	7.13	17.41	.000
TASK_ORD BY SIM_QUES	.54	2	.27	.66	.519
LRN GOAL BY SIM_QUES	3.29	2	1.65	4.02	.021
TASK_ORD BY LRN GOAL BY SIM_QUES	.51	2	.26	.63	.536

Within groups comparisons between the three questions, data collapsed over task order

Control task subjects

Overall vs underlying	Z = -1.5724	2-Tailed P =	.1159
Overall vs strategic	Z = -1.1847	2-Tailed P =	.2361
underlying vs strategic	Z = -.2801	2-Tailed P =	.7794

pattern search subjects

Overall vs underlying	Z = -3.8230	2-Tailed P =	.0001
Overall vs strategic	Z = -3.6214	2-Tailed P =	.0003
underlying vs strategic	Z = -1.0142	2-Tailed P =	.3105

EXPERIMENT 6b

Learning trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	116.92	22	5.31		
TASK	.08	1	.08	.02	.901

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	191.25	22	8.69		
TRL BLOCK	75.00	1	75.00	8.63	.008
TASK BY TRL BLOCK	6.75	1	6.75	.78	.388

test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	535.83	44	12.18		
TASK	2.04	1	2.04	.17	.684
LRN GOAL	92.04	1	92.04	7.56	.009
TASK BY LRN GOAL	7.04	1	7.04	.58	.451

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	223.00	44	5.07		
TRL BLOCK	337.50	1	337.50	66.59	.000
TASK BY TRL BLOCK	.67	1	.67	.13	.719
LRN GOAL BY TRL BLOCK	2.67	1	2.67	.53	.472
TASK BY LRN GOAL BY TRL BLOCK	.17	1	.17	.03	.857

transfer between learning and test trials

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	224.46	22	10.20		
TASK	7.52	1	7.52	.74	.400

Tests involving 'TRL BLOCK' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	113.29	22	5.15		
TRL BLOCK	111.02	1	111.02	21.56	.000
TASK BY TRL BLOCK	.19	1	.19	.04	.850

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	282.25	22	12.83		
TASK	6.75	1	6.75	.53	.476

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	174.92	22	7.95		
PHASE SCORE	3.00	1	3.00	.38	.545
TASK BY PHASE SCORE	10.08	1	10.08	1.27	.272

prediction questions, analysis done on data in Figure 6.7

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	88166.63	43	2050.39		
TASK	387.66	1	387.66	.19	.666
LRN GOAL	43770.58	1	43770.58	21.35	.000
TASK BY LRN GOAL	811.04	1	811.04	.40	.533

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	54836.29	86	637.63		
QUE_TYPE	6517.81	2	3258.90	5.11	.008
TASK BY QUE_TYPE	863.90	2	431.95	.68	.511
LRN GOAL BY QUE_TYPE	5974.19	2	2987.09	4.68	.012
TASK BY LRN GOAL BY QUE_TYPE	1488.92	2	744.46	1.17	.316

Within group comparisons

control task subjects

Old-wrong vs Old-correct	Z = -.3360	2-Tailed P =	.7369
Old-wrong vs new	Z = -3.1284	2-Tailed P =	.0018
Old-correct vs new	Z = -3.4576	2-Tailed P =	.0005

pattern search subjects

Old-wrong vs Old-correct	Z = -1.2545	2-Tailed P =	.2097
Old-wrong vs new	Z = -.4080	2-Tailed P =	.6832
Old-correct vs new	Z = -.9799	2-Tailed P =	.3271

Prediction questions, analysis done on data where situations defined by last two elements only.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	84568.33	42	2013.53		
TASK_ORD	549.91	1	549.91	.27	.604
LRN_GOAL	44137.35	1	44137.35	21.92	.000
TASK_ORD BY LRN_GOAL	920.57	1	920.57	.46	.503

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	52473.40	84	624.68		
QUE_TYPE	9021.30	2	4510.65	7.22	.001
TASK_ORD BY QUE_TYPE	923.74	2	461.87	.74	.480
LRN_GOAL BY QUE_TYPE	5179.05	2	2589.53	4.15	.019
TASK_ORD BY LRN_GOAL BY QUE_TYPE	444.24	2	222.12	.36	.702

Within group comparisons

control task subjects

Old-wrong vs Old-correct	Z =	-.2427	2-Tailed P =	.8083
Old-wrong vs new	Z =	-3.1480	2-Tailed P =	.0016
Old-correct vs new	Z =	-3.8472	2-Tailed P =	.0001

pattern search subjects

Old-wrong vs Old-correct	Z =	-1.0342	2-Tailed P =	.3011
Old-wrong vs new	Z =	-.0628	2-Tailed P =	.9499
Old-correct vs new	Z =	-.9941	2-Tailed P =	.3202

Spearman rank correlation coefficients between total learning, test, and transfer phase scores and total prediction questions.

Pi = person interaction task, F= factory task , psc = pattern search, ctk= control task

	pi-psc	f-psc	pi-ctk	f-ctk
TOT_LRN			-.2758	-.5163
			N(12)	N(11)
			Sig .386	Sig .104
TOT_TST	.7811	.8211	-.0958	.1264
	N(11)	N(12)	N(12)	N(11)
	Sig .005	Sig .001	Sig .767	Sig .711
TOT_TRA	.7778	.6608	.0991	.1227
	N(11)	N(12)	N(12)	N(11)
	Sig .005	Sig .019	Sig .759	Sig .719
	TOT_QUE2	TOT_QUE2	TOT_QUE2	TOT_QUE2

prediction questions done on data where overlap between learning and test phase is removed - so Old-wrong is not actually learning phase Old-correct and vice versa

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	86015.86	43	2000.37		
TASK_ORD	265.47	1	265.47	.13	.717
LRN_GOAL	45586.14	1	45586.14	22.79	.000
TASK_ORD BY LRN_GOAL	951.78	1	951.78	.48	.494

- -

Tests involving 'QUE_TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	56284.75	86	654.47		
QUE_TYPE	6545.00	2	3272.50	5.00	.009
TASK_ORD BY QUE_TYPE	667.53	2	333.77	.51	.602
LRN_GOAL BY QUE_TYPE	5828.14	2	2914.07	4.45	.014
TASK_ORD BY LRN_GOAL BY QUE_TYPE	1826.88	2	913.44	1.40	.253

Within group comparisons

control task subjects

Old-wrong vs Old-correct	Z =	-.6160	2-Tailed P =	.5379
Old-wrong vs new	Z =	-3.1243	2-Tailed P =	.0018
Old-correct vs new	Z =	-3.3277	2-Tailed P =	.0009

pattern search subjects

Old-wrong vs Old-correct	Z =	-1.3255	2-Tailed P =	.1850
Old-wrong vs new	Z =	-.2795	2-Tailed P =	.7798
Old-correct vs new	Z =	-1.0017	2-Tailed P =	.3165

Spearman rank correlation coefficients between total learning, test, and transfer phase scores and total prediction questions.

Pi = person interaction task, F= factory task , psc = pattern search, ctk= control task

	pi-psc	f-psc	pi-ctk	f-ctk
TOT_LRN	.	.	-.1317	-.4098
	N(12)	N(12)	N(12)	N(11)
	Sig .	Sig .	Sig .683	Sig .211
TOT_TST	.6773	.8594	-.0603	.1057
	N(12)	N(12)	N(12)	N(11)
	Sig .016	Sig .000	Sig .852	Sig .757
TOT_TRA	.6726	.7201	-.0071	.0532
	N(12)	N(12)	N(12)	N(11)
	Sig .017	Sig .008	Sig .983	Sig .876
	TOT_QUE3	TOT_QUE3	TOT_QUE3	TOT_QUE3

transfer between tasks

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	2539.54	44	57.72		
TASK_ORD	.84	1	.84	.01	.904
LRN GOAL	943.76	1	943.76	16.35	.000
TASK_ORD BY LRN GOAL	8.76	1	8.76	.15	.699

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	331.21	44	7.53		
PHASE SCORE	44.01	1	44.01	5.85	.020
TASK_ORD BY PHASE SCORE	3.76	1	3.76	.50	.483
LRN GOAL BY PHASE SCORE	133.01	1	133.01	17.67	.000
TASK_ORD BY LRN GOAL BY PHASE SCORE	5.51	1	5.51	.73	.397

within groups comparisons comparing overall performance between transfer and test phase with data collapsed over task

For the pattern search group t(22) = -1.28, p = 0.214

For the control task group t(22) = 4.7, p = 0.000

similarity questions

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	120.81	44	2.75		
TASK_ORD	3.06	1	3.06	1.12	.297
LRN GOAL	76.56	1	76.56	27.89	.000
TASK_ORD BY LRN GOAL	5.06	1	5.06	1.84	.181

Tests involving 'SIM_QUE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	23.28	88	.26		
SIM_QUE	17.68	2	8.84	33.42	.000
TASK_ORD BY SIM_QUE	.29	2	.15	.55	.578
LRN GOAL BY SIM_QUE	1.29	2	.65	2.44	.093
TASK_ORD BY LRN GOAL BY SIM_QUE	.12	2	.06	.24	.790

Within groups comparisons between the three questions, data collapsed over task order

Control task subjects			
Overall vs underlying	Z = -2.2713	2-Tailed P =	.0231
Overall vs strategic	Z = -2.8304	2-Tailed P =	.0046
underlying vs strategic	Z = -1.6773	2-Tailed P =	.0935
pattern search subjects			
Overall vs underlying	Z = -3.8230	2-Tailed P =	.0001
Overall vs strategic	Z = -3.6214	2-Tailed P =	.0003
underlying vs strategic	Z = -.5916	2-Tailed P =	.5541

EXPERIMENT 6a vs EXPERIMENT 6b (SIMPLIFIED VERSION OF THE FACTORY TASK VERSUS THE OLD VERSION OF THE FACTORY TASK)

Between group comparisons for those with a control task subjects, simplified learning goal vs normal learning goal. Any prediction question variable with a 2 in it comes from the second analysis of prediction question data (described above). Any prediction question variable with a 3 in it comes from the third analysis of prediction question data (described above).

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_LRN	12.14229	140.72727	12.14229	6.70130	1.81193	.193
FHLF_LRN	.10672	75.54545	.10672	3.59740	.02967	.865
LHLF_LRN	20.75132	142.55303	20.75132	6.78824	3.05695	.095
TOT_TST	5.31258	133.64394	5.31258	6.36400	.83479	.371
FHLF_TST	.55369	99.09848	.55369	4.71898	.11733	.735
LHLF_TST	2.43610	117.30303	2.43610	5.58586	.43612	.516
OW1	702.10804	23640.4040	702.10804	1125.73353	.62369	.439
OC1	6.00296	12902.2096	6.00296	614.39093	.00977	.922
NEW1	180.36891	6984.84848	180.36891	332.61183	.54228	.470
TOT_QUE1	245.52296	7249.71467	245.52296	345.22451	.71120	.409
OW2	514.65744	23351.7677	514.65744	1111.98894	.46283	.504
OC2	171.54242	15759.6170	171.54242	750.45795	.22858	.638
NEW2	316.33729	6557.57576	316.33729	312.26551	1.01304	.326
OW3	584.10280	23554.0614	584.10280	1121.62197	.52077	.478
OC3	42.68775	13412.6263	42.68775	638.69649	.06684	.799
TOT_QUE3	216.40326	7895.51300	216.40326	375.97681	.57558	.456

Between group comparisons for those with a pattern search subjects, simplified learning goal vs normal learning goal. Any prediction question variable with a 2 in it comes from the second analysis of prediction question data (described above). Any prediction question variable with a 3 in it comes from the third analysis of prediction question data (described above).

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
TOT_TST	24.72859	914.57576	24.72859	43.55123	.56780	.459
FHLF_TST	10.67194	204.54545	10.67194	9.74026	1.09565	.307
LHLF_TST	2.91041	322.39394	2.91041	15.35209	.18958	.668
OW1	.66700	23102.7146	.66700	1100.12927	.00061	.981
OC1	291.04084	22722.7273	291.04084	1082.03463	.26898	.609
NEW1	134.91436	27030.3030	134.91436	1287.15729	.10482	.749
TOT_QUE1	76.78222	14932.3171	76.78222	711.06272	.10798	.746
OW2	5.56653	25027.5253	5.56653	1191.78692	.00467	.946
OC2	416.87253	21413.3207	416.87253	1019.68194	.40883	.529
NEW2	297.26614	30701.7677	297.26614	1461.98894	.20333	.657
OW3	8.96739	23211.8056	8.96739	1105.32407	.00811	.929
OC3	346.05567	23053.2197	346.05567	1097.77237	.31523	.580
TOT_QUE3	122.56258	14258.2211	122.56258	678.96291	.18051	.675

transfer between tasks

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	4171.46	88	47.40		
TASK_ORD	37.63	1	37.63	.79	.375
FAC_TASK	121.92	1	121.92	2.57	.112
LRN GOAL	1457.51	1	1457.51	30.75	.000
TASK_ORD BY FAC_TASK	23.38	1	23.38	.49	.484
TASK_ORD BY LRN GOAL	2.30	1	2.30	.05	.826
FAC_TASK BY LRN GOAL	27.76	1	27.76	.59	.446
TASK_ORD BY FAC_TASK BY LRN GOAL	32.51	1	32.51	.69	.410

Tests involving 'PHASE SCORE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	727.79	88	8.27		
PHASE SCORE	128.38	1	128.38	15.52	.000
TASK_ORD BY PHASE SCORE	9.63	1	9.63	1.16	.283
FAC_TASK BY PHASE SCORE	3.80	1	3.80	.46	.500
LRN GOAL BY PHASE SCORE	131.67	1	131.67	15.92	.000
TASK_ORD BY FAC_TASK BY PHASE SCORE	34.17	1	34.17	4.13	.045
TASK_ORD BY LRN GOAL BY PHASE SCORE	43.13	1	43.13	5.22	.025
FAC_TASK BY LRN GOAL BY PHASE SCORE	23.38	1	23.38	2.83	.096
TASK_ORD BY FAC_TASK BY LRN GOAL BY PHASE SCORE	10.55	1	10.55	1.28	.262

similarity questions

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	267.08	88	3.04		
TASK_ORD	9.03	1	9.03	2.98	.088
FAC_TASK	.03	1	.03	.01	.919
LRN GOAL	116.28	1	116.28	38.31	.000
TASK_ORD BY FAC_TASK	.28	1	.28	.09	.762
TASK_ORD BY LRN GOAL	1.84	1	1.84	.61	.439
FAC_TASK BY LRN GOAL	2.53	1	2.53	.83	.364
TASK_ORD BY FAC_TASK BY LRN GOAL	3.34	1	3.34	1.10	.297

Tests involving 'SIM_QUE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	59.33	176	.34		
SIM_QUE	31.69	2	15.85	47.01	.000
TASK_ORD BY SIM_QUE	.75	2	.38	1.11	.331
FAC_TASK BY SIM_QUE	.25	2	.13	.37	.691
LRN GOAL BY SIM_QUE	4.00	2	2.00	5.93	.003
TASK_ORD BY FAC_TASK BY SIM_QUE	.08	2	.04	.12	.884
TASK_ORD BY LRN GOAL BY SIM_QUE	.11	2	.06	.16	.848
FAC_TASK BY LRN GOAL BY SIM_QUE	.58	2	.29	.87	.423
TASK_ORD BY FAC_TASK BY LRN GOAL BY SIM_QUE	.53	2	.26	.78	.459

The general questions asked were:

The Pattern Question - "Could you try to describe what sort of pattern you thought Clegg was using to respond to your behaviour".

The Control Question - "How did you get Clegg to behave as you wanted him to?".

Answers categorised as No information or Wrong

Subject **IS**.

Answer to the Pattern Question - "No idea what so ever - seemed almost random".

Answer to the Control Question - "I didn't get him to behave as I wanted him to! However I did try out a few different strategies but always seemed to be wrong or loose track."

Subject **JL**.

Answer to the Pattern Question - "Didn't find any recurring pattern. At times I thought I had found a pattern but Clegg then made a response which didn't fit. (1) If I was Affectionate - I thought a certain response would occur. (2) Tried to over compensate at times, began by trying to match Clegg's responses - i.e. moving up & down the scale as I thought Clegg was.

Answers categorised as Partially Correct

Subject **MT**.

Answer to the Pattern Question - No answer given.

Answer to the Control Question - " When Clegg was being friendly or too rude then it seemed to be best to work him down in stages, i.e. Very friendly - Friendly - Polite etc. Rather than trying to make a single jump to counter his mood. It was hard to keep him constantly at one level. Being Very Rude made him Very Rude.

Subject **DJH**.

Answer to the Pattern Question - "If I was ruder than him he would then get ruder, if I was politer than him he would then get politer. Towards the Very Rude and Loving ends of the scale this became more pronounced. However there were times when this appeared not to work!"

Answers categorised as Correct

Subject **GG.**

Answer to the Pattern Question - "Clegg would respond to the direction in which my behaviour moved by moving in the same direction, only further. The amount by which his response would exceed the change in my behaviour was approximately double the distance of the change I had made. If my change of behaviour was extreme his change, within the limits of the scale, would be almost twice as extreme."

Answer to the Control Question - "By Moving from the position of Clegg's behaviour halfway towards the desired state of behaviour. Then trying to keep my changes small or indeed by remaining at the desired state of behaviour."

Subject **SSC**.

Answer to the Pattern Question - "If Clegg was Loving and I was Very Friendly then he would go Polite; if I went 2 below him then he would go 2 below that, if I went 3 below him he would go 3 below that, if I went 4 below him then he would go 4 below that, etc. etc. It wouldn't happen like this <u>every</u> time."

Answer to the Control Question - "By following the pattern."

The 'describing' pattern search subjects simply gave verbalisations literally describing what was on the interactions. This occurred irrespective of learning goal, therefore only the example from one of the goal groups is given.

'describe' pattern search subject

"....Clegg is polite, I am going to be polite, Clegg is now very polite I am going to be very rude, Clegg is now very rude I am going to be very friendly...."

'explain' pattern search subject

".... He is indifferent, I said very friendly he goes to loving, I said very friendly he goes to very polite. I shall type in very friendly to see how far up he goes. I think what is happening is that he is closing in on very friendly, so if I type in very friendly he should go to very friendly - nope but it is close. If I try friendly - he goes to very friendly. If I try very rude, Clegg is now very rude. If I type in very friendly he should go all the way to loving. - yes If I go to indifferent he should go all the way to very rude yes...."

'explain' pattern search subject

".... Lets go right down to very rude to see if he behaves the same way at the bottom of the scale as he does at the top. As I thought he is following me, now if I move up to rude he should go one above me if he is behaving as at the top of the scale. Yup, If I move up 2 he has moved from rude to friendly. That is a similar relationship from the top part of the scale... If I have moved up 3 steps he has moved up three steps plus a bit more. Now if I move down to very rude....I think the relationship is that if I move up 1 he will move up 1 further than me, If I move up 2 he will move up 2 further than me maybe plus one more. I don't know about 3, lets see what happens with three..."

'explain' control task group

"....Clegg is loving so I think I will be ruder to him to reduce his affection. Clegg has been very rude. If I am nicer to him it may move him up towards the polite area I am trying to get him to....Now he is very rude again, I might be polite to him, I don't want to encourage him. Perhaps I will be polite. I guess I just have to be very friendly. Clegg is loving. I'll try being rude. Again I haven't tried to be cool - He is polite, good. I think I should stick being polite towards him. He has been encouraged and is very polite, may be I should slow him down and be indifferent...."

'explain' control task group

"....He is now Very rude, he probably thought I was being pretentious. He is now very rude. Oh no! He is swinging quite a lot. I will continue trying polite. Try being indifferent. Looks like I've got to go and meet him. I will try being affectionate. I will try being very friendly. I will continue being friendly. Last time he was very friendly I was polite he got very angry. I will continue being polite. He is indifferent I will try being polite hopefully he will be polite again...."

'explain' dual goal group

"....Clegg is no loving. I want to get him to be polite, lets see If I am cool he is very rude. I wonder if it is a continuous cycle. Will it change later. Clegg is very friendly, I too am very friendly, I will be very polite, Clegg is very polite, I will be very cool Clegg is now rude, I want to get him to be polite. I will see what happens if I do not encourage him in his affectionate ways. I will be very friendly. Lets see, I have to get him to be polite...."

List and explanation of contents of floppy disc

The floppy disc accompanying this thesis contains 6 SPSS files (compatible with version SPSS 6.0 or 6.1 for windows). Each file contains the information that produced the statistics connected with each experiment.

File name	The experiment it is relevant to
lrn_gl_ef.sav	Experiment 1 : learning goal effect experiment
observe.sav	Experiment 2 : observe experiment
Con_verb.sav	Experiment 3 : concurrent verbalisation experiment
RNG.sav	Experiment 4 : random number generation experiment
memory.sav	Experiment 5 : memory experiment
trans6a_6b.sav	Experiments 6a & 6b : transfer experiments

The variables within each file are all self explanatory and are illustrated in the variable label (double click on the variable name at the head of the column and then click on variable label). It should be noted that there are nearly always three different sets of variables for the prediction questions with the suffixes 1,2, & 3. The variables with a '1' after them are the data that are used in the graphs reported in the experiments. Those with a '2' after them are the data that was used for the second analysis where the data was refined defining prediction situations by the last two elements (see pg. 31 for a further explanation of this analysis). Those with a '3' after them are the data that was used for the third analysis where the data was refined removing the Old-wrong and Old-correct prediction questions that were of an opposite situation type in the learning phase (see pg. 220 for a further explanation of this analysis).

Bibliography

BIBLIOGRAPHY

Anderson, J.R. (1983). *The architecture of cognition.* Cambridge, Mass.: Harvard University Press.

Anderson, J.R. (1987). Skill acquisition: Compilation of weak method problem solutions. *Psychological Review, 94(2).,* 192-210.

Baddeley, A. & Hitch, G.J. (1974). Working Memory. In G. Bower (Ed.*), Recent advances in learning and motivation* (Vol. 8, pp47-90). New York: Academic Press.

Baddeley, A. (1986). *Working Memory.* Oxford: Oxford University Press.

Baddeley, A. (1990). *Human Memory - theory and practise.* Lawrence Erlbaum Associates, Hove (UK).

Baddeley, A. (1992). Is working memory working? The fifteenth Bartlett lecture. *Quarterly Journal of Experimental Psychology,* 44A, 1-31.

Barclay, J.R., (1973). The abstraction of linguistic ideas. *Cognitive Psychology.* 4, 229-254.

Bassok, M., & Holyoak, K.J., 1993. Pragmatic knowledge and conceptual structure: Determinants of Transfer between quantitative domains. In D.K., Detterman & R.J. Sternberg (Eds.), *Transfer on trial: intelligence, cognition, and instruction.* Alex Publishing Corporation: Norwood, New Jersey.

Bassok, M., Wu, L-L. & Olseth, K.L. (1995). Judging a book by its cover: Interpretive effects of content on problem-solving transfer. *Memory and Cognition, 23*, 354-367.

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

Berry, D.C. & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79*, 251-272.

Berry, D.C. (1991). The role of action in implicit learning. *Quarterly Journal of Experimental Psychology, 43A,* 881-906.

Berry, D.C. (1994) . A step too far? *Behavioural and Brain Sciences, 17*, 397-398.

Berry, D.C., & Dienes, Z. (1993). *Implicit Learning: Theoretical and empirical issues.* Lawrence Erlbaum Associates, Hove (UK).

Bielaczyc, K., Pirolli, P.L., & Brown, A.L. (1995) Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*, 221-252.

Bourke, P.A., Duncan, J., & Nimmo-Smith, I., (1996). A general factor involved in dual task performance decrement. *Quarterly Journal of Experimental Psychology,* 49A (3), 525-545.

Bovair, S., Kieras, D.E. & Polson, P.G. (1990). The acquisition and performance of text-editing skill: A cognitive complexity analysis. *Human Computer Interaction, 5*, 1-48.

Broadbent, D. E., Fitzgerald, P., & Broadbent M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology, 77*, 33-50.

Brown, A.L. (1990). Domain-Specific Principles Affect Learning and Transfer in Children. Cognitive Science, 14, 107-134.

Buchner, A., Funke, J., & 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.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989) Self-explanations: How students study and use examples in learning to solve problems, *Cognitive Science, 13*, 145-182.

Chi, M.T.H., de Leeuw, N., Chiu, M-H. & LaVancher, C. (1994) Eliciting self-explanations improves learning. *Cognitive Science, 18,* 439-478.

Cleeremans, A. (1993*). Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing.* MIT Press.

Cleeremans, A. (1994). *Behavioural and Brain Sciences, 17*, 402-403.

Cohen, A., Ivry, R.I., & Keele, S.W. (1990). Attention and structure of event sequences. *Journal of Experimental Psychology: General,* 120, 235-253.

Detterman, D.K. (1993) The Case for the Prosecution: Transfer as an Epiphenomenon. Knowledge and conceptual Structure: Determinants of Transfer Between Quantitative Domains. In Douglas K. Detterman & Robert J. Sternberg (Eds.) *Transfer on Trial: Intelligence, Cognition and Instruction.* Norwood, New Jersey: Ablex Publishing Corporation.

Dienes, Z. & Perner, J. (1994). Dissociable definitions of consciousness. *Behavioural and Brain Sciences, 17*, 403-404.

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

Dienes, Z., Broadbent D., & Berry, D.C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal Of Experimental Psychology: Learning, Memory And Cognition.* 17, 875-887.

Dulaney, D.E., Carlson, R.A. & Dewey, G.I. (1984). A case of syntactical learning and judgement: How conscious and how abstract? *Journal of Experimental Psychology: General, 113*, 541-555.

Evans, F.J. (1978), Monitoring attention deployment by random number generation: An index to measure subjective randomness. Bulletin of the Psychonomic Society. Vol 12 (1), 35-38

Evans, J.St.B.T. & Over, D. (1997). Rationality and reasoning: the problem of deductive competence. *Cahiers de Psychologie Cognitive*, 16, 3-38.

Ferguson, G.A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology, 10*, 121-131.

Geddes, B.W., & Stevenson, R.J. (in press). Learning goals and the implicit and explicit learning distinction. *Quarterly Journal of Experimental Psychology,*

Gick, M. L. & Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355.

Gick, M. L. & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 125*, 1-38.

Gilhooly, K.J., Logie, R.H., Wetherick, N.E. & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. Memory & Cognition, 21, 115-124.

Green, R.E.A., & Shanks, D.R. (1993). On the existence of independent explicit and implicit learning systems: An examination of some evidence. *Memory and Cognition, 21*, 304-317.

Hayes, J.R., & Simon, H.A, (1974). Understanding written problem instructions, in L.W. Gregg (Ed.), *Knowledge and Cognition*. Lawrence Erlbaum Associates, Hillsdale NJ.

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

Hebb, D.O. (1949). *The Organisation of Behaviour.* New York: Wiley.

Holyoak, K.J. & Gattis, M. (1994). *Behavioural and Brain Sciences, 17*, 406-407.

Johnson-Laird, P. N., (1983). Mental models in cognitive science. *Cognitive Science*, 4, 71-115.

Judd, C.H. (1908). The relation of special training to general intelligence. *Educational Review, 36*, 28-42.

Katz, I.R. (1991). Assessing transfer of a complex skill. In *Proceedings of the 13th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Keele, S.W. & Jennings, P.J. (1992). Attention in the representation of sequence: Experiment and theory. *Human Movement Science*, 11, 125-138.

Kessler, C.M. (1988). Transfer of programming skills in novice LISP learners. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh, PA.

Kieras, D.E. & Bovair, S. (1986). The acquisition of procedures from text: A production-system analysis of transfer of training. *Journal of Memory and Language, 25*, 507-524.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12, 1-55.*

Komatsu, L.K., (1992). Recent Views of Conceptual Structure. *Psychological Bulletin*, 112, 3, 500-526.

Lebowitz, M. (1986). Integrated learning: Controlling explanation. *Cognitive Science, 10*, 219-240.

Logan, D. & Klapp, S.T. (1992). Automatizing alphabetic arithmetic: 1. Is extended practise necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 179-195.

Logan, G.D. (1988). Toward and instance theory of automatization. *Psychological Review, 95*, 492-527.

Logie, R.H. (1991). Characteristics of visual short-term memory. *European Journal of Cognitive Psychology*. 1, 275-284.

Logie, R.H., Zucco G., & Baddeley A. (1991). Interference with visual short-term memory. *Acta Psychologica*, 75, 55-74.

Marescaux, P-J., Luc, F., & Karnas, G. (1989). Modes d'apprentissage selectif et nonselectif et connaissances acquies au control d'un process: Evaluation d'un modele simule. *Cahiers de Psychologie Cognitive, 9,* 239-264.

Mathews, R. C. (1991). The forgetting algorithm: How fragmentary knowledge of exemplars can abstract knowledge. *Journal of Experimental Psychology: General, 120,* 117-119.

Mathews, R.C., Buss, R.R., Stanley, W.B., Blanchard-Fields, F., Cho, J.R. & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 1083-1100.

Matthews. R. (1990). Abstractivenss of implicit grammar knowledge. Comments on Perruchet and Pacteau's analysis of synthetic grammar learning. *Journal of Experimental Psychology: General, 119*, 412-416.

Mawer, R.F. & Sweller, J. (1982). Effects of subgoal density and location on learning during problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 252-259.

McKendree, J. & Anderson, J.R. (1987). Effect of practise on knowledge and use of basic LISP. In J.M. Carroll (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. Cambridge, MA: MIT Press.

Medin, D.L., & Schafer, M.M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Michalski, R. S., (1989). Two-tiered concept meaning, inferential matching, and conceptual cohesiveness. In S. Vosniadou & A. Ortony (Eds.)., *Similarity and analogical reasoning* (pp. 122-145). Cambridge, England: Cambridge University Press.

Murphy, G. L., & Medin, D.L., (1985). The role of theories on conceptual coherence. *Psychological Review*, 92, 289-316.

Neves, D.M. & Anderson, J.R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J.R. Anderson (Ed.), *Cognitive Skills and Their Acquisition.* Hillsdale, NJ: Erlbaum.

Newell, A., & Simon, H.A. (1972). *Human Problem Solving.* Engelwood Cliffs, NJ: Prentice-Hall.

Ng. E. & Bereiter, C. (1995). Three Levels of Goal Orientation in Learning. In A. Ram & D.B. Leake (Eds.) *Goal-Driven Learning.* Cambridge, Mass., London England, MIT Press.

Nosofsky, R.M., Clark, S.E, Shin, H.J., (1989). Rules and exemplars in categorisation, identification and recognition. *Journal Of Experimental Psychology: Learning, Memory And Cognition. 15,2,*282-304.

Novick, L.R. (1990). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14,* 510-520.

Owen E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology, 77,* 272-284.

Pennington, N., Nicolich, R. & Rahm, J. (1995). Transfer of training between cognitive subskills: Is Knowledge Use Specific? *Cognitive Psychology, 28*, 175-224.

Perruchet, P. & Gallego, J. (1994). *Behavioural and Brain Sciences, 17*, 415-416

Perruchet, P. & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule extraction or explicit fragmentary knowledge? . *Journal of Experimental Psychology: General, 119,* 264-275.

Perruchet, P. (1992). Learning from complex rule-governed environments: On the proper function of conscious and unconscious processes. In Umilta C. and Moscovitch M. (Eds.), *Attention and Performance XV.* Cambridge, Mass, and London: MIT Press.

Perruchet, P., Gallego, J. & Pacteau, C. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in experimental situations. *Cognitive Psychology, 22*, 493-516.

Pirolli, P., & Bielaczyc, K, (1989). Empirical analyses of self-explanation and transfer in learning to program. In *Proceedings of the 11th annual Conference of the Cognitive Science Society* (pp.450-457). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pirolli, P., & Recker, M, (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction*, 12, 235-275.

Polson, P.G. & Kieras, D.E. (1985). A qualitative model of learning and performance of text-editing knowledge/ In C. Borman and B. Curtis (Eds.), Proceedings of the CIII 1985

Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery.

Porter, D. (1986). *A functional examination of intermediate cognitive processes.* Unpublished DPhil thesis, University of Oxford.

Porter, D. (1988). Computer games and human performance. In *Proceedings of 11th Symposium on Psychology in the Department of Defence.* Colorado Springs: US Airforce Academy, 251-255.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115-119.

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 88-94.

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An essay on the Cognitive Unconscious.* Oxford University Press, New York.

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor G. W. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 492-502.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour, 5,* 855-863.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219-235.

Reber, A.S., Walkenfeld, F.F., & Hernstadt, R., (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: 77*, 317-327.

Reed, S.K, Ernst, G.W. & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology*, 6, 436-450.

Rips, L.J., (1989b). Similarity, typicality, and categorisation. In S. Vosniadou & A. Ortony (Eds.)., *Similarity and analogical reasoning* (pp. 21-59). Cambridge, England: Cambridge University Press.

Sanderson, P. (1990). *Implicit and explicit control of a dynamic task: Empirical and conceptual issues.* EPRL Report 90-02, University of Illinois. Referred to in Berry & Dienes ,1993, pp 28,35.

Schoenfield, A.H. (1985). *Mathematical Problem Solving.* Orlando: Academic Press.

Scholtz, J. & Wiedenbeck, S. (1990). Learning second and subsequent programming languages: A problem of transfer. *International Journal of Human-Computer-Interaction. 2*, 51-72.

Seger, C. A. (1994). Implicit Learning. *Psychological Bulletin*, 115, 163-196.

Servan-Schrieber, E. & Anderson, J.R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 592-608.

Shanks, D. R., & St. John, M. F., (1994). Characteristics of dissociable human learning systems. *Behavioural and Brain Sciences,* 17, 367-447.

Simon, H. A. & 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.

Singley, M.K., and Anderson, J.R. (1989) *The Transfer of Cognitive Skill.* Cambridge, Massachusetts, and London, England: Harvard University Press

Singley, M.K., and Anderson, J.R. (1989) *The transfer of text-editing skill. International Journal of Man-Machine Studies. 22*, 403-423.

Smith, E. E., & Sloman, S.A., (1994). Similarity versus rule-based categorisation. *Memory and Cognition*, 22, 377-386.

Squire, L.M., & Frambach, M., (1990). Cognitive skill learning on amnesia. Psychobiology 18(1), 109-117.

Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalisation, instruction and practise in a simulated process control task. *Quarterly Journal of Experimental Psychology, 41A*, 553-577.

Stevenson, R.J. & Palmer, J.A., (1994). Learning: Principles, processes and practises. Cassel.

Stevenson, R.J. (1997). Deductive reasoning and the distinction between implicit and explicit processes: A commentary on Evans and Over. *Cahiers de Psychologie Cognitive*, 16, 222-229.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Sweller, J., Mawer, R.F. & Ward, M.R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112*, 639-661.

Thorndike, E.L. & Woodworth, R.S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8,* 247-261.

Thorndike, E.L. (1913). mental discipline in high school studies. *Journal of Educational Psychology, 15*, 1-22.

VanLehn, K. & Jones, R.M. (1993) Learning by explaining examples to oneself: A computational model. In S. Chipman and A.L. Meyrowitz (Eds.) *Foundations of Knowledge Acquisition.* Kluwer Academic Publishers.

Vollmeyer, R., & Burns, B.D. (1995). Does hypothesis-instruction improve learning? in J.D. Moore & J.F. Lehman (Eds.) Proceedings of the Seventeenth Annual conference of the Cognitive Science Society (pp. 771-776). Mahwah, NJ; Hove, UK: LEA.

Vollmeyer, R., Burns, B.D., & Holyoak, K.J. (1996) The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75-100.

Whittlesea, B.W.A., & Dorken, M.D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General, 122*, 227-248.

Wisniewski, E.J. & Medin, D.L. (1995). Harpoons and Long Sticks: The Interaction of Theory and Similarity in Rule Induction. In A. Ram & D.B. Leake (Eds.) *Goal-Driven Learning*. Cambridge, Mass., London England, MIT Press.

Wu, Q. & Anderson, J.R. (1991). Knowledge transfer among programming languages. In Proceedings of the 13th Annual Meeting of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.