Information selection and belief updating in hypothesis evaluation

Feeney, Aidan

http://hdl.handle.net/10026.1/344
http://dx.doi.org/10.24382/3843
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.
Theory
I am what is around me

Women understand this.
One is not duchess
A hundred yards from a carriage.

These then are portraits:
A black vestibule;
A high bed sheltered by curtains.

These are merely instances.

Wallace Stevens
Collected Poems, 1955
This thesis is concerned with the factors underlying both selection and use of evidence in the testing of hypotheses. The work it describes examines the role played in hypothesis evaluation by background knowledge about the probability of events in the environment as well as the influence of more general constraints.

Experiments on information choice showed that subjects were sensitive both to explicitly presented probabilistic information and to the likelihood of evidence with regard to background beliefs. It is argued - in contrast with other views in the literature - that subjects' choice of evidence to test hypotheses is rational allowing for certain constraints on subjects' cognitive representations. The majority of experiments in this thesis, however, are focused on the issue of how the information which subjects receive when testing hypotheses affects their beliefs. A major finding is that receipt of early information creates expectations which influence the response to later information. This typically produces a recency effect in which presenting strong evidence after weak evidence affects beliefs more than if the same evidence is presented in the opposite order. These findings run contrary to the view of the belief revision process which is prevalent in the literature in which it is generally assumed that the effects of successive pieces of information are independent. The experiments reported here also provide evidence that processes of selective attention influence evidence interpretation: subjects tend to focus on the most informative part of the evidence and may switch focus from one part of the evidence to another as the task progresses. In some cases, such changes of attention can eliminate the recency effect.

In summary, the present research provides new evidence about the role of background beliefs, expectations and cognitive constraints in the selection and use of information to test hypotheses. Several new findings emerge which require revision to current accounts of information integration in the belief revision literature.
# TABLE OF CONTENTS

## CHAPTER 1 - SELECTING EVIDENCE TO EVALUATE HYPOTHESES: THE ROLE OF BACKGROUND KNOWLEDGE

1.1 Introduction ............................................................................................................. 1  
   1.1.1 Introduction to the thesis ............................................................................... 1  
   1.1.2 Introduction to Chapter 1 .............................................................................. 2  

1.2. Wason's selection task ............................................................................................ 3  
   1.2.1 The abstract selection task ............................................................................ 3  
   1.2.2 The thematic selection task: Early research .................................................. 7  

1.3 Recent developments - the influence of pragmatics ............................................. 9  
   1.3.1 A Not So Recent Development - Pragmatic Reasoning Schemas ............... 9  
   1.3.3 The Relevance Theoretic Approach .............................................................. 15  

1.4. Subjective probability and the selection task ....................................................... 21  
   1.4.1. Oaksford and Chater's Rational Analysis ..................................................... 21  
   1.4.3 The effect of subjective probability .............................................................. 30  

1.5 An alternative selection task ................................................................................... 35  
   1.5.1 The pseudodiagnosticity paradigm: An alternative selection task ............... 35  
   1.5.2. The Pseudodiagnosticity task: A closer inspection ...................................... 36  

1.6 Summary and conclusions ...................................................................................... 40  

## CHAPTER 2 - EXPERIMENTS 1 AND 2: SUBJECTIVE PROBABILITY AND THE SELECTION OF EVIDENCE FOR THE EVALUATION OF HYPOTHESES

2.1 Introduction ............................................................................................................. 45  

2.2 Experiment 1  
   2.2.1 Introduction ................................................................................................... 48  
   2.2.1.1 Manipulating the probability of the features .............................................. 48  
   2.2.1.2 The use of diagnostic and pseudodiagnostic information ......................... 49  
   2.2.2. Method ......................................................................................................... 51  
   2.2.3. Results and Interpretation ............................................................................ 56  
   2.2.3.1 Information Choice .................................................................................... 56  
   2.2.3.2 Confidence Ratings .................................................................................... 60
2.2.3.2. a Confidence ratings with probabilistic evidence ...................................... 61
2.2.3.2. b Confidence ratings with propositional evidence ..................................... 66
2.2.4 Discussion ..................................................................................................... 69

2.3. Experiment 2
2.3.1 Introduction ................................................................................................... 72
2.3.2 Pre-test .......................................................................................................... 74
2.3.3 Method .......................................................................................................... 76
2.3.4. Results .......................................................................................................... 82
2.3.4.1. Evidence selections ................................................................................... 82
2.3.4.2. Usefulness Ratings .................................................................................... 83
2.3.5 Discussion ..................................................................................................... 87

2.4. General discussion .......................................................................................... 89

2.5 Summary .......................................................................................................... 97

CHAPTER 3 - THE ORDER EFFECTS PARADIGM AND MODELS OF BELIEF REVISION

3.1. Introduction ........................................................................................................ 99

3.2. Order effects and information integration ........................................................ 100
  3.2.1. The order effects paradigm .......................................................................... 100
  3.2.2. Three Explanations for Primacy Effects ...................................................... 103

3.3. Models of belief revision as information integration .......................................... 109
  3.3.1. Cognitive Algebraic Models ........................................................................ 109
  3.3.2. Procedural Models ....................................................................................... 115
    3.3.2.1. How the evidence is encoded ................................................................. 118
    3.3.2.2 How the evidence is processed ............................................................... 122
    3.3.2.3 How the adjustment is accomplished ...................................................... 124
  3.3.3. A Problem for Existing Models ................................................................... 126
    3.3.3.1 Inference, domain-specific knowledge structures, and information integration .............................................................. 128
    3.3.3.2 Background knowledge about correlated features ................................. 133
    3.3.3.3 The importance of pragmatics ............................................................... 136
      3.3.3.3.1 Pragmatics and meaning ................................................................. 136
      3.3.3.3.2 Pragmatic determinants of evidential importance ............................ 139
3.4. Summary and conclusions ..................................................................................... 141
  3.4.1 Summary ....................................................................................................... 141
  3.4.2 Conclusions ................................................................................................... 143

CHAPTER 4 - EXPERIMENTS 3, 4 AND 5: ORDER EFFECTS, EXPECTATION
AND BELIEF REVISION

4.1. Introduction ............................................................................................................ 146
  4.1.1 Another look at Experiment 1 ....................................................................... 146
  4.1.2 Examining the role of expectation using the order effects paradigm .......... 150

4.2. Experiment 3
  4.2.1 Introduction ................................................................................................... 151
  4.2.2 Method .......................................................................................................... 152
  4.2.3 Results ........................................................................................................... 157
    4.2.3.1 Confidence Ratings .................................................................................... 157
    4.2.3.2 Direction of Revision ................................................................................. 159
    4.2.3.3 Expectations about the second piece of evidence ...................................... 159
  4.2.4 Discussion ..................................................................................................... 161

4.3 Experiment 4
  4.3.1 Introduction ................................................................................................... 163
  4.3.2 Pre-test .......................................................................................................... 168
    4.3.2.1 Establishing the base rates ......................................................................... 168
    4.3.2.2 Establishing the level of perceived associations amongst the features ..... 169
  4.3.3 Experiment 4 - Method ................................................................................. 172
  4.3.4 Results ........................................................................................................... 175
    4.3.4.1 Confidence ratings ..................................................................................... 175
    4.3.4.2 Expectations about the second piece of evidence ...................................... 178
  4.3.5 Discussion ..................................................................................................... 180

4.4 Experiment 5
  4.4.1 Introduction ................................................................................................... 182
  4.4.2 Method .......................................................................................................... 186
  4.4.3 Results ........................................................................................................... 188
    4.4.3.1 Confidence measures .................................................................................. 188
    4.4.3.2 Expectation measures ............................................................................... 192
  4.4.4 Discussion ..................................................................................................... 194

4.5 General discussion ................................................................................................... 197
4.6 Summary .................................................................................................................. 201

CHAPTER 5 - EXPERIMENTS 6, 7 AND 8: SOURCES OF EVIDENTIAL STRENGTH

5.1 Introduction ............................................................................................................. 203

5.2 Experiment 6
  5.2.1 Introduction ................................................................................................... 205
  5.2.2 Method .......................................................................................................... 211
  5.2.3 Results ........................................................................................................... 217
    5.2.3.1 Confidence measures ................................................................................. 217
    5.2.3.2 Expectation measures ................................................................................. 219
  5.2.4 Discussion ..................................................................................................... 221

5.3 Experiment 7
  5.3.1 Introduction ................................................................................................... 225
  5.3.2 Method .......................................................................................................... 227
  5.3.3 Results ........................................................................................................... 231
  5.3.3 Discussion ..................................................................................................... 234

5.4 Experiment 8
  5.4.1 Introduction ................................................................................................... 235
  5.4.2 Method .......................................................................................................... 240
  5.4.3 Results ........................................................................................................... 244
  5.4.4 Discussion ..................................................................................................... 251

5.5 General discussion ................................................................................................... 256

5.6 Summary .................................................................................................................. 262

CHAPTER 6 - OVERVIEW

6.1 Introduction ............................................................................................................. 266

6.2 A summary and initial discussion of the experiments on information selection .................................................. 266
6.3 A summary and initial discussion of the experiments on information use ........ 273
   6.3.1 Experiments 3, 4 and 5: The role of expectation in belief revision .......... 273
   6.3.2 Experiments 6, 7 and 8: Assigning values to evidence .......................... 279

6.4 The role of background knowledge in information selection and belief revision ......................................................................................................................... 285

6.4 What do information selection and information use have in common? .......... 292
   6.4.1 Information selection as decision making ............................................... 293
   6.4.2 Belief revision, inference and relevance .................................................. 295

6.5 Concluding remarks ....................................................................................... 298

APPENDICES

Appendix 1 - Example stimulus materials from Experiments 1 - 8 ................. 300

Appendix 2 - Summary tables, tables of means and frequency tables from
Experiments 1 - 4 ........................................................................................................ 322

Appendix 3 - Summary tables from the analyses of subjects' confidence ratings and
expectations in Experiments 5 - 8 ............................................................................. 331

Appendix 4 - Stimulus materials, mean ratings and standard deviations from the
pre-tests .................................................................................................................... 336

REFERENCES ........................................................................................................ 341
LIST OF TABLES

Table 1.1  
Predicted expected information gain versus actual proportion of cards selected from Oaksford and Chater's (1994) meta-analysis of results on the standard abstract selection task. ........................................................................................................26

Table 2.1  
Features used in Experiment 1. The 2nd feature column refers to the feature kept constant across all conditions. ........................................................................................................54

Table 2.2  
Independent and Dependent variables involved in Experiment 1. 56

Table 2.3  
Response frequencies, broken down by Rarity and Likelihood, for Experiment 1........................................................................................................57

Table 2.4  
Initial evidence, remaining possible evidence, possible outcomes, and the results of those outcomes when subjects are given propositional information on the pseudodiagnosticity task. 59

Table 2.5  
Means, standard deviations and number of subjects involved in the significant two way interaction between Likelihood and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1. 63

Table 2.6  
The confidence ratings of subjects who chose diagnostic evidence in Experiment 1, broken down by Likelihood and Feedback, and the corresponding Bayesian probabilities given the evidence. 63

Table 2.7  
The means, standard deviations and subject numbers involved in the interaction between Rarity and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1. 64

Table 2.8  
Means, standard deviations and subject numbers involved in the significant three way interaction between Rarity, Feedback and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1. 65

Table 2.9  
Problem contents and mean frequency estimations for the features used in Experiment 2. 76

Table 2.10  
Problem contents and features used in Experiment 2........................................................................................................81

Table 2.11  
The independent and dependent variables involved in Experiment 2. 82

Table 2.12  
Percentages of subjects choosing each item broken down by instructional manipulation and feature type for Experiment 2. 83
Table 2.13  
Mean item selections broken down by features and instructional manipulation from Experiment 2 ................................................................. 83

Table 2.14  
Means, standard deviations and significant differences between means for the interaction between Hint and Item Rated from the analysis of subjects' usefulness ratings in Experiment 2 ........................................................................ 85

Table 2.15  
The means and standard deviations involved in the interaction between Rarity, Hint, and the Item Rated from the analysis of subjects' usefulness ratings in Experiment 2. ................................................................. 86

Table 2.16  
The probabilities and utilities associated with each possible outcome from the task used in Experiment 2 ....................................................................................... 91

Table 2.17  
Expected information yield for each possible piece of information, broken down by Probability, from Experiment 2 ................................................................. 94

Table 4.1  
The independent and dependent variables involved in Experiment 3 .......... 156

Table 4.2  
Means and standard deviations from the analysis of subjects confidence ratings in Experiment 3. ................................................................. 157

Table 4.3  
Means and standard deviations involved in the significant interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 3. ................................................................. 158

Table 4.4  
Number (and equivalent percentages in parentheses) of subjects, broken down by Order and problem content, who revised their initial confidence ratings upwards, downwards, or left them unchanged in Experiment 3 .............. 159

Table 4.5  
The means and standard deviations involved in the significant interaction between Order and problem content from the analysis of subjects' expectations in Experiment 3. ................................................................. 160

Table 4.6  
The estimated base rates of the features selected for inclusion in Experiment 4 ....................................................................................... 168

Table 4.7  
The rate at which objects possessing the rare feature were expected to possess each of the common features (standard deviations in parentheses) from the pre-test for Experiment 4 ....................................................................................... 171

Table 4.8  
Combinations of high and low association features used in Experiment 4 broken down by whether the feature formed part of a strong or weak piece of evidence. ....................................................................................... 174

Table 4.9  
Independent and dependent variables from Experiment 4. .............................. 175
Table 4.10
Means and standard deviations from the analysis of subjects' confidence ratings in Experiment 4. ................................................................. 176

Table 4.11
Means and standard deviations involved in the significant interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 4. ....................................................... 177

Table 4.12
The means and standard deviations involved in the significant interaction between Association, Order and Before/After from the analysis of subjects' confidence ratings in Experiment 4. ................................................................. 177

Table 4.13
Means and standard deviations from the analysis of subjects' expectations in Experiment 4. ................................................................. 178

Table 4.14
Means and standard deviations involved in the interaction between Order and problem content from the analysis of subjects' expectations in Experiment 4. ................................................................. 180

Table 4.15
Independent and dependent variables from Experiment 5. ................................. 188

Table 4.16
Means and standard deviations from the analysis of subjects' confidence ratings in Experiment 5. ................................................................. 189

Table 4.17
Means, standard deviations, and significant differences between means for the interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 5. ................................................................. 189

Table 4.18
Means, standard deviations and significant differences between means for the interaction between Likelihood and Order from the analysis of subjects' confidence ratings from Experiment 5. ................................................................. 191

Table 4.19
Means, standard deviations and significant differences between means for the interaction between Likelihood, Order and Before/After from the analysis of subjects' confidence ratings in Experiment 5. ................................................................. 191

Table 4.20
Means and standard deviations from the analysis of subjects' confidence ratings in Experiment 5. ................................................................. 192

Table 4.21
Means, standard deviations and significant differences between means for the interaction of Order with problem content from the analysis of subjects' expectations in Experiment 5. ................................................................. 193

Table 5.1
Perceived rarity and associations for the features used in Experiment 6. ............ 215

Table 5.2
The independent and dependent variables involved in Experiment 6. ............ 216
Table 5.3
Means and standard deviations from the analysis of subjects' confidence ratings in Experiment 6. ................................................................. 217

Table 5.4
Means, standard deviations and significant differences between means for the interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 6. ................................................................. 218

Table 5.5
Means, and standard deviations from the analysis of subjects' expectations in Experiment 6. ................................................................. 220

Table 5.6
Means, standard deviations and significant differences between means involved in the interaction between Order and problem content from the analysis of subjects' expectations in Experiment 6. ................................................................. 221

Table 5.7
The independent and dependent variables involved in Experiment 7.............. 231

Table 5.8
Means and standard deviations from the analysis of subjects' confidence ratings in Experiment 7. ................................................................. 232

Table 5.9
Means, standard deviations, and significant differences between means for the interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 7. ................................................................. 233

Table 5.10
Means and standard deviations from the analysis of subjects' expectations in Experiment 7. ................................................................. 234

Table 5.11
The independent and dependent variables involved in Experiment 8.............. 244

Table 5.12
Means and standard deviations from the analysis of subjects' confidence ratings in Experiment 8. ................................................................. 245

Table 5.13
Means, and standard deviations, involved in the two-way interaction between Order and Rarity and in the two-way interaction between Order and Before/After. Both from the analysis of subjects' confidence ratings in Experiment 8. ................................................................. 246

Table 5.14
Means, standard deviations and significant differences between means from the two way interaction between Likelihood Ratio and Before/After from the analysis of subjects' confidence ratings in Experiment 8. ................................................................. 248

Table 5.15
Means, standard deviations, and significant differences between means for the three-way interaction between Order, Before/After, and Rarity from the analysis of subjects' confidence ratings in Experiment 8. ................................................................. 249
Table 5.16
Means, standard deviations, and significant differences between means for the three way interaction between Order, Before/After, and Likelihood Ratio from the analysis of subjects' confidence ratings in Experiment 8.

LIST OF FIGURES

Figure 1.1
The abstract version of the Wason selection task.

Figure 1.2
Mynatt et al's (1993) conceptualisation of the pseudodiagnosticity task.

Figure 2.1
The percentages of subjects choosing diagnostic and pseudodiagnostic information in Experiment 1 broken down by the level at which the first feature was said to be present amongst instances of X.

Figure 2.2
The significant two way interaction between Likelihood and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1.

Figure 2.3
The significant three way interaction between Rarity, Feedback and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1.

Figure 2.4
The interaction between the Item Rated and problem content from the analysis of subjects' usefulness ratings in Experiment 2.

Figure 2.5
The interaction between Hint and Item Rated from the analysis of subjects' usefulness ratings in Experiment 2.

Figure 2.6
The interaction between Rarity, Hint and Item Rated from the analysis of subjects' usefulness ratings in Experiment 2.

Figure 2.7
The structure of the variant of the pseudodiagnosticity paradigm used in Experiment 2.

Figure 3.1
Diagrammatic representation of the Shanteau and Nagy conceptualisation of Information Integration.

Figure 4.1
The significant three way interaction between Probability, Feedback and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1.

Figure 4.2
The significant two-way interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 3.
Figure 4.3
The significant interaction between Order and problem content from the analysis if subjects' expectations in Experiment 3 ................................................................. 160

Figure 4.4
Interaction of Order with Before/After from the analysis of subjects' confidence ratings in Experiment 4........................................................................................................ 176

Figure 4.5
The significant interaction between Association, Order and Before/After from the analysis of subjects' confidence ratings in Experiment 4. .......................... 178

Figure 4.6
The significant interaction between Order and problem content from the analysis of subjects' expectations in Experiment 4. ...................................................... 179

Figure 4.7
The significant interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 5.......................................................... 190

Figure 4.8
The interaction between Likelihood and Order from the analysis of subjects' confidence ratings in Experiment 5.................................................................................. 190

Figure 4.9
The interaction between Likelihood, Order and Before/After from the analysis of subjects' confidence ratings in Experiment 5......................................................... 192

Figure 4.10
The interaction between Order and problem content from the analysis of subjects' expectations in Experiment 5.................................................................................. 193

Figure 5.1
The significant interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 6............................................................. 218

Figure 5.2
The interaction between Order, Rarity, and Before/After from the analysis of subjects' confidence ratings in Experiment 6...................................................................... 219

Figure 5.3
The interaction between Order and problem content from the analysis of subjects' expectations in Experiment 6.................................................................................. 220

Figure 5.4
A comparison of the interaction between Order and Before/After from Experiment 7 with the three-way interaction between Order, Before/After, and Rarity from Experiment 6 .................................................................................. 233

Figure 5.5
Massaro and Friedman's (1990) general stage model of information integration.............................................................................................................................. 239

Figure 5.6
The interaction between Order and Rarity from the analysis of subjects' confidence ratings in Experiment 8.......................................................... 247
Figure 5.7
The interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 8. ................................................................. 247

Figure 5.8
The interaction between Likelihood Ratio and Before/After from Experiment 8. ............................................................................................................. 248

Figure 5.9
The interaction between Order, Rarity, and Before/After from the analysis of subjects' confidence ratings in Experiment 8. .................................................. 249

Figure 5.10
The interaction between Order, Likelihood Ratio, and Before/After from the analysis of subjects' confidence ratings in Experiment 8. ........................................ 250
Acknowledgements

This thesis would not have been embarked upon without the inspiration of Ruth Byrne at Trinity College, Dublin. Neither could it have been attempted without the financial support of the Faculty of Human Sciences at the University of Plymouth.

My director of studies, Jonathan Evans, has, at all times, been encouraging, helpful, and humane. Many thanks are due also to John Clibbens, my second supervisor. Four other people in the psychology department here at Plymouth also deserve profuse thanks: Kenny Coventry, for (sometimes not so gently) forcing me to make leaps forward; Mike Tucker and Simon Handley for their comradeship; and Pete Bradon for helping me to prop up many unstable bars.

My parents, Seamus and Frances, have been phoning up almost every Sunday night for the last three years making sure that I have never felt like an emigrant. Finally, most thanks must go to Aine who doesn't realise how immeasurably better my life has been since she became part of it.
Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This work was financed with the aid of a studentship from the Faculty of Human Sciences at the University of Plymouth.

A programme of advanced study was undertaken, which included a postgraduate course on the philosophy and psychology of science.

Relevant scientific conferences were regularly attended at which work was presented.

Conferences attended and papers presented:


Order effects and belief updating in pseudodiagnostic reasoning. Paper presented at the BPS Cognitive Section XII Annual Conference, University of Bristol, 7 - 9 September, 1995.

Information order, expectation and pragmatics: The case of sequential pseudodiagnostic evidence. Poster presented at the BPS Cognitive Section, 3rd International Conference on Thinking, University College London, 29 - 31 August, 1996

Selecting evidence to evaluate hypotheses: Pragmatically determined focus or maximised expected information gain? Paper presented at the IX Conference of the European Society for Cognitive Psychology, University of Wurzburg, 4 - 8 September, 1996.

Signed .................................

Date ................................

I. T

Aqq
CHAPTER 1 - SELECTING EVIDENCE TO EVALUATE HYPOTHESES: THE ROLE OF BACKGROUND KNOWLEDGE

1.1 INTRODUCTION

1.1.1 Introduction to the thesis

For any state of affairs in the world there are almost always at least two possible explanations, causes, or descriptions in which one is likely to have different degrees of confidence. The basic questions which this thesis asks are:

- How do we go about choosing information to evaluate how likely each of these possibilities is?

- Once that information has been received how do we use it in revising our beliefs about what is actually the case?

Typically these questions are asked, and answered, separately, in wholly different ways, by different groups of researchers. It is obvious, however, that they are very closely related. One of the goals of this thesis will be to demonstrate how it may be possible to answer these questions in the same terms.

This thesis is not unique in attempting to draw together the traditionally separate questions of information choice and information use. For example, Evans, Manktelow and Over (1993), and Johnson-Laird (1994) have sketched proposals for the unification of the fields of reasoning and decision-making. Evans et al attempted to do so by reconceiving a problem used in the reasoning literature as a decision making task, whilst Johnson-Laird attempted to apply a theoretical framework, induced from work on deductive reasoning, to the problem of probabilistic thinking. This work is interesting because it attempts to integrate two separate traditions and thereby hints at the possibility of a unified theory of thinking. Although this thesis does not have an aim as grand as a unified theory of high-
level cognition, it is hoped that the reader will find it to contain some interesting, and well
justified, experiments. It is also hoped that the sense which has been made of their results,
and the attempts made to integrate these results with the existing literature, will make
sense to the reader.

The central theme of the thesis is how background knowledge is used to inform
both the selection, and use, of information for the evaluation of hypotheses. Accordingly,
the first chapter will be a review of the literature on the effects of background knowledge
on the selection of information for hypothesis testing. This will be followed by a chapter
detailing the results of two experiments investigating some sources of information in
background knowledge and their effect on information selection. Next, the literature on
belief revision will be critically reviewed. The subsequent three chapters will describe a
series of experiments carried out in the light of that review. The final chapter will relate the
experimental findings to each other, and to the existing literature.

1.1.2 Introduction to Chapter 1

This, the first chapter of the thesis, will concern itself with previous work on the
role of background knowledge in the selection of information to test hypotheses. By so
doing, it will pave the way for the experimental work to be described in Chapter 2. In a
sense, the literature on information selection has been task driven. Most of the
experimental work to be described in this chapter centres around Wason's Selection Task
(Wason, 1966). Recently, however, some very interesting theoretical developments have
taken place. These have been the attempts made to explain the effects of subjective
probability (e.g. Oaksford and Chater, 1994) and pragmatics (e.g. Sperber, Cara and
Girotto, 1995; Evans and Over, 1996b) on subjects' behaviour in the task. The first section
of this chapter will, therefore, be taken up with a description of early work using both
abstract and thematic versions of the Wason Selection Task, whilst subsequent sections
will outline the recent theoretical developments. There will be three such developments
discussed. The first of these is pragmatic reasoning schemas (Cheng and Holyoak, 1985).
It will be argued that whilst the introduction of a schema-based approach revolutionised
the theoretical landscape, such an approach is inadequate in accounting for the effects of background knowledge on subjects' information selection. The next two developments to be discussed will be the relevance theoretic approach (Sperber, Cara, and Girotto, 1995) and Oaksford and Chater's model of maximised expected information gain. Once again, it will be argued that, either of these approaches, on their own, are inadequate. However, it will be concluded that the work of Evans and Over (1996) offers an indication of how these approaches might be integrated in some way, in order to produce a satisfactory account of the part played by background knowledge in the selection of information for the evaluation of hypotheses.

A superficial reading of the literature on information selection for hypothesis evaluation might lead the casual observer to conclude that there exists only one task for the study of how people select evidence. Although such a conclusion would be understandable, it would also be unjustified. There exists a suitable, but under-utilised, alternative. This alternative is the pseudodiagnosticity paradigm (Doherty, Mynatt, Tweney and Schiavo, 1979). A later section of this chapter will present and, using some of the ideas discussed in the previous sections, analyse the pseudodiagnosticity paradigm. It is hoped to demonstrate not only that the pseudodiagnosticity paradigm is a suitable tool for experimentation, but also that it offers a relatively easy means of obtaining evidence about the factors influencing information selection for hypothesis evaluation. As the section on recent theoretical developments will show, such new evidence is badly needed, for there is a debate currently raging in the literature which cries out for arbitration.

1.2. WASON'S SELECTION TASK

1.2.1 The abstract selection task

There are a variety of tasks in the literature where subjects are asked to select some information in order to verify the truth or falsity of a rule or to decide between alternatives.
The most widely researched of these is the Wason selection task (Wason, 1966). The task, in its original form is presented in Figure 1.1.

![Figure 1.1: The abstract version of the Wason selection task](image)

Subjects are given a conditional rule of the form if p then q. They are then shown one side of each of four cards corresponding to p, not-p, q, and not-q. Finally, subjects are asked which cards must be turned over in order to decide whether the rule is true or false. The normatively correct choice is to select the p card (A in Figure 1.1) and the not-q card (7 in Figure 1.1). These are the only two cards which might falsify the rule. The not-p and q cards are of no use because the rule does not specify any relationship between instances of not-p and q. Therefore, the truth or falsity of the rule cannot be inferred from a not-p card with a q on the back or from a q card with a not-p on the back.

The task seems simple. It is deceptively so. In Wason's original paper only 10% of subjects made the correct selections. This result has been replicated several times in the literature (e.g. Wason and Johnson-Laird, 1970; Jackson and Griggs, 1988). However, on the many so-called concrete versions of the selection task - where the conditional rule specifies a relationship with which subjects are assumed to be familiar - normatively correct card selections can rise to 100% (see Yachanin, 1986), and regularly rise above 75%. It is with these versions of the selection task that this section will be primarily concerned. However, we will first consider some of the factors which are thought to affect card selections on the abstract form of the task.
Evans, Newstead, and Byrne (1993) identify two main contemporary theories of subjects' performance on the abstract selection task. These are the heuristic/analytic approach of Evans (1984; 1989) and the mental models account offered by Johnson-Laird and Byrne (1991). The Evans account of performance on the task rests on his distinction between a pre-attentive stage of processing and a later analytic stage. The former stage determines what cards the subject attends to whilst any reasoning which might take place occurs in the latter stage. Evans holds that in the case of the abstract selection task only the pre-attentive stage is involved. He also argues that the cards attended are determined by their linguistic relevance. He claims that there are two determinants of relevance at work on the task. These have been termed the NOT-heuristic and the IF-heuristic (Oaksford and Stenning, 1992). The word if in the rule directs attention to the circumstance it hypothesises whilst the word not directs attention to the circumstance it negates.

Although there is some empirical support for this account in the literature (see Evans, Ball, and Brooks, 1987; Evans, 1992; Evans, 1996) it has also come in for criticism of several types. The first of these is that Evans' account suggests that subjects do not engage in any reasoning on the task, thereby deeming them to be irrational (Johnson-Laird and Byrne, 1991; Oaksford and Stenning, 1992). For the purposes of this review this is not a central criticism, although it is of greater importance in the rationality debate where philosophers of mind such as Stich (1990) have cited work on the selection task as evidence for the view that humans are inherently irrational.

Three further criticisms of the theory have also been made. Firstly Johnson-Laird and Byrne point out that it does not specify a deductive device which is necessary to account for the numerous cases of human deductive competence in the literature. Secondly, it has been claimed (Oaksford and Chater, 1993) that the Evans' account is poorly specified and does not lead to clear experimental predictions. However, such a criticism seems unfair in the light of recent work by Evans, Clibbens, and Rood (1996) which actually tested, and confirmed, empirical predictions derived from Evans' account. Finally, Oaksford and Chater (1993) claim that it is unlikely that an appropriate set of heuristics will be found to
overcome the computational intractability problems faced by all current theories of deduction.

In response to the deductive mechanism criticism mentioned at the start of the previous paragraph, Evans (1993) has outlined how his approach could adopt the mental models approach of Johnson-Laird and Byrne (1991). Their approach to deductive reasoning is based on the idea that it depends on constructing a model (or models) of the situation described by the premises. Next, a conclusion must be derived which is true in the models and which makes explicit something which was only implicit in the premises. Finally, the validity of the conclusion is checked by searching for alternative models consistent with the premises in which the initial conclusion is false. If no counterexample is found then the conclusion is valid. It is this final stage which constitutes Johnson-Laird and Byrne's criterion for the attribution of rationality (see Johnson-Laird and Byrne, 1993 for a discussion of this point).

As applied to the abstract selection task, mental model theory accounts for the data by claiming that there are certain principles governing what subjects will include in their initial representation of a proposition. For example, they account for matching bias by claiming that subjects will represent positive items. Thus they will represent the A and the 3 in an initial single model of the rule contained in Figure 1. In a case where the rule was "If there is not an A on one side of the card then there is a 3 on the other", Johnson-Laird and Byrne suggest that subjects will initially represent the positive items (A and 3) in two separate models.

Whilst not wanting to go into the details of their account (for detailed examinations see Evans, Newstead, and Byrne, 1993; Evans, 1991;1992) it is worth pointing out that Johnson-Laird and Byrne claim that anything which leads subjects to include not-q in their initial representation of the problem will facilitate their performance on the task. Evans, Newstead and Byrne point out that this fact leads to the conclusion that the position of Johnson-Laird and Byrne is very similar to that of Evans. This cannot be the case as the former authors claim that subjects do engage in some explicit reasoning on the task whilst
the latter author does not. However, they do have in common the claim that factors to do with the linguistic structure of the problem are crucial in determining subjects' card selections. From the point of view of this review what is important is the claim that certain factors contained in the problem make certain cards relevant (for a definition of Relevance see Section 1.3.3) thus causing them to be either (à la Evans) chosen or (à la Johnson-Laird and Byrne) represented in an initial model. In other words, both accounts stress the importance of focus to card selection.

What the literature on the abstract selection task demonstrates is the importance of non-logical factors to subjects' information choices in hypothesis evaluation. The importance of such factors is highlighted by work done with concrete versions of the selection task. This latter work also emphasises the importance of background knowledge in the explanation of subjects' behaviour. The next section will review this literature and accordingly our theme will initially switch from linguistically determined focus to the effects of background knowledge. However, the notion of linguistic relevance will be briefly returned to in the concluding section of the chapter.

1.2.2 The thematic selection task: Early research

Although much work continues to be done with the abstract selection task, in recent years work on thematic versions of the task has achieved pre-eminence. This work started in the early 1970s with the publication by Wason and Shapiro (1971) of the first paper on the topic and continues apace. The basic issue is one of determining how background knowledge is used by subjects in choosing information for the evaluation of hypotheses. As we shall see, there are many possible ways in which this might happen and there are also many different factors which contribute to the process. This section will concentrate on the work done in the 1970s and early 1980s using the thematic selection task. The reader should bear in mind throughout the rest of this review that the work on the role of background knowledge in the selection of information for the evaluation of hypotheses which will be discussed, is essential to an understanding of the experiments to be presented in Chapter 2.
As mentioned above, the original demonstration of a facilitatory effect for thematic materials was reported by Wason and Shapiro (1971). These authors substituted the rule "Every time I go to Manchester I travel by car" for the standard abstract rule described in the previous section. Subjects were then shown cards which had "Manchester" (p), "Leeds" (not-p), "car" (q), and "train" (not-q) printed on their exposed sides. Subjects were asked to decide which cards they would turn over in order to tell whether the experimenter's rule was true or false. Significantly more subjects chose the appropriate combination of cards than did subjects in a control condition who received a standard abstract form of the task.

Although facilitation with this rule was replicated several times in the 1970s (e.g. Gilhooly and Falconer, 1974), later attempts at replication failed (Mankteelow and Evans, 1979; Griggs and Cox, 1982; Reich and Ruth, 1982). These later studies, combined with Mankteelow and Evans' (1979) failure to obtain facilitation with other thematic materials and the population specific facilitation produced by Johnson-Laird, Legrenzi, and Legrenzi's (1972) postal rule, led Mankteelow and Evans to suggest that perhaps thematic versions of the selection task were a test of memory rather than a test of reasoning. Results from studies using the postal rule materials were central to this line of argument.

The postal rule is "If a letter is sealed then it has a 50 lire stamp on it". Subjects are shown a display of four envelopes, two of which are face up and two of which are face down. Of the face up envelopes, one has a 50 lire stamp and the other has a 40 lire stamp. Of the other two envelopes one is sealed whilst the other is unsealed. Johnson-Laird et al found that the vast majority of subjects given these materials (81%) selected the correct cards. However, Mankteelow and Evans argued that the rate of facilitation was this high only because there was a postal rule in operation in England at this time which meant that sealed letters had to have a stamp of higher value than did unsealed letters. Griggs and Cox (1982) found that if these materials were given to American subjects in the absence of a context there was no effect of facilitation. Likewise, Golding (1981) found that amongst English subjects, only the performance of those who remembered the original postal rule (it had been abolished in England shortly after the study of Johnson-Laird et al) was facilitated.
As pointed out by Evans, Newstead and Byrne (1993), there was an awareness amongst researchers at the time that the "memory cueing" hypothesis (and Pollard's, 1982 similar "availability" explanation) were unsatisfactory before Cheng and Holyoak (1985) took issue with the claim that subjects were making selections on the task based on their memory of domain-specific experiences. That subjects were not selecting based on specific remembered experiences was already strongly suggested in work by D'Andrade (1982) who found facilitation for materials with which subjects could have had no prior experience. Equally bad news for the memory cueing/availability hypotheses was the fact that Cox and Griggs (1982) had found significant transfer between thematic materials which did not produce facilitation and materials which did. This only happened when subjects were given the facilitatory materials first. This strongly suggested that there was some reasoning by analogy going on in the study for which a simple memory cueing approach could not account. Although there has been considerable debate since then about transfer from facilitatory to non-facilitatory materials (see Berry, 1983; Klaczynski, Gelfand, and Reese, 1989), nevertheless the stage was set for the emergence of a new account. It is this account and the work which it has inspired which will next be discussed.

1.3 RECENT DEVELOPMENTS - THE INFLUENCE OF PRAGMATICS

1.3.1 A Not So Recent Development - Pragmatic Reasoning Schemas

In this section of the chapter, Cheng and Holyoak's (1985) notion of pragmatic reasoning schemas will be discussed at some length. Although this discussion is not directly relevant to the studies which will be presented in Chapter 2, it is necessary for at least two reasons. The first of these is that Cheng and Holyoak's (1985) work was to become central to the development of the field in the ensuing years. It seems to have triggered an explosion of papers (both theoretical and experimental) on the selection task. Secondly, the pragmatic reasoning schemas approach to the selection task contrasts strongly with the approaches of both Sperber, Cara and Girotto (1995) and Oaksford and Chater (1994). Accordingly, an exposition of Cheng and Holyoak's approach will provide a useful background against which to assess more recent theoretical work.
As mentioned above, Cheng and Holyoak (1985) took issue with the then approach to thematic facilitation on the selection task. Specifically, they pointed to D'Andrade's results and the fact that Manktelow and Evans (1979) had been unable to obtain facilitation using materials with which subjects should have had experience. Instead of a simple "selection based on remembered instances model" they proposed an approach based on "a type of knowledge structure qualitatively different from those postulated by other theories of deductive reasoning" (pg. 395). This knowledge structure is the schema, first proposed by Kant (1787/1963) as a means of organising our experience of the world. The notion of schemas as organising framework has also been used by Bartlett (1932) and Piaget (1967) and became prominent in the 1970s with Schank's (1972) work on "primitive acts", Minsky's (1975) work on "frames" in artificial intelligence, and Schank and Abelson's (1977) proposal of "scripts" to account for people's stereotypical knowledge of frequently experienced situations.

Schemas, generally speaking, have the following features (Eysenck and Keane, 1990):

- **relations** which can take simple (hit, kick etc.) or complex (enable, cause etc.) forms
- **slots** which take variables or other sub-schemata
- **values** which refer to the specific items which fill slots
- the property of **general application** so that they may be applied in many specific situations

Cheng and Holyoak postulated the existence of a number of reasoning schemas which they claimed were "*induced from ordinary life experiences such as 'permissions' 'obligations' and 'causations'*" (pg. 395). Further to the properties listed above, pragmatic reasoning schemas were claimed to be defined in terms of goals and relationships to these goals. It was also claimed that the primary importance of goals constrained subjects to pragmatically useful inferences.
The initial case study offered by Cheng and Holyoak was that of the permission schema which they defined as describing "a type of regulation in which taking a particular action requires satisfaction of a particular precondition" (pg. 396). The permission schema may be characterised by the following four rules:

1. If the action is to be taken, then the precondition must be satisfied
2. If the action is not to be taken, then the precondition need not be satisfied
3. If the precondition is satisfied, then the action may be taken
4. If the precondition is not satisfied, then the action must not be taken

Cheng and Holyoak claimed that most of the rules used to demonstrate thematic facilitation on the selection task fit a permission schema. As an illustration of how this might work we will consider one of the rules most frequently used on thematic versions of the selection task. The Drinking Age rule (Griggs and Cox, 1982) has been used to demonstrate the facilitating effect of thematic materials many times (e.g. Griggs and Cox, 1983; Griggs, 1984). The rule states that "If a person is drinking beer then the person must be over 19 years of age". In this case p (drinking beer) is the action that is to be taken and q (being over 19 years of age) is the precondition that must be satisfied. The rules for the permission schema tell the subject that p (Rule 1) and not-q (Rule 4) must be chosen as these are the only rules which contain the deontic imperative must. Given the experimental rule, subjects avoid the selection of the q card because Rule 3 merely states that the action may be carried out if the precondition is satisfied. It is not important to check drinkers who are over 19 in this case.

This type of analysis explains most of the thematic materials research which had been carried out up until 1985. It explains why Wason and Shapiro's (1971) original Towns and Transport rule is unreliable (it is not a permission rule) and the effectiveness of the Postal Rule scenario amongst subjects who have had experience of a similar rule. It also explains the importance of the context in which the thematic rule is placed. For example, the usual context for the Drinking Age rule is that of the subject imagining herself to be a police officer checking to see whether people are conforming to certain rules. Evans and
Pollard (1987) have found that the Drinking Age rule is not an effective facilitator when it is administered without this context. Cheng and Holyoak (1985) found that when the Postal Rule, which Griggs and Cox (1982) had found to be ineffective with American subjects, was administered in the context of maximising revenue for the postal service, the performance of American subjects was just as good as that of subjects from Hong Kong where a similar rule had been in operation.

In addition to the result just mentioned, Cheng and Holyoak provided several other experimental findings which suggested the importance of pragmatic reasoning schemas to performance on the thematic selection task. In Experiment 2 of their paper they presented an abstract selection task where subjects were told that they were an authority checking regulations all of which were of the form "If one is to take action 'A' then one must fulfil precondition 'P'." The fours cards used were as follows:

- has taken action A (p)
- has not taken action A (not-p)
- has fulfilled precondition P (q)
- has not fulfilled precondition P (not-q)

These materials obviously contain no thematic content, yet Cheng and Holyoak predicted that they would evoke the permission schema. This they did with 61% of subjects choosing the p and not-q cards compared with 19% correct choices in a standard abstract control condition. A third experiment showed that when subjects were asked to rephrase permission statements the words must or may were used more often than they were when subjects were asked to rephrase arbitrary rules. Additional support for pragmatic reasoning schemas comes from Girotto and co-workers (e.g. Girotto, Light, and Colbourn, 1988; Light, Blaye, Gilly, and Girotto, 1990) who have demonstrated the facilitation effects of permission schemas on the cards selected by children as young as 10 years old.

Although there have been extensions of the original work on pragmatic reasoning schemas (see Cheng, Holyoak, Nisbett, and Oliver, 1986; Kroger, Cheng, and Holyoak, 1993; Politzer, and Nguyen-Xuan, 1992; Holyoak and Cheng, 1995), the approach has been
criticised almost since it first appeared. Apart from the methodological complaints of Jackson and Griggs (1990) and Griggs and Cox (1992) (see Evans et al, 1993 for a discussion of this debate) there have been several theoretical attacks on the approach.

One serious attack on the pragmatic reasoning schema approach to information choice on the selection task has come from those investigators (Manktelow and Over, 1991; Gigerenzer and Hug, 1992) who have demonstrated that subjects' information selection on thematic versions of the selection task can be reversed if they are asked to take a different perspective. Recently, Holyoak and Cheng (1995) have enlarged somewhat upon their initial position in order to account for these findings. However, the remainder of this section will focus on other problems which have been pointed out along with Cheng and Holyoak's replies to these various criticisms. These other problems are more central to the concerns of this chapter as they relate to more general questions of how information is organised in background knowledge. As shall be seen in subsequent sections of this chapter, a recent theoretical focus in the literature has been on how best to characterise the information stored in long-term memory which subjects bring to bear on hypothesis evaluation tasks.

Apart from the perspective shift findings, two of the most common criticisms of the schema approach are that it is firstly too narrow and secondly too inflexible. As Johnson-Laird and Byrne (1995) have pointed out, there exists no support for the existence of pragmatic reasoning schemas outside the experimental literature on the selection task. This narrow field of focus is not in itself disastrous for the approach. However, before schema theory is to be widely accepted, a large body of work, using a wide variety of tasks, must be built up. Similarly, Oaksford and Chater (1993), drawing on Fodor's (1983) work on modularity have commented that Cheng and Holyoak's assumption that the knowledge drawn upon by subjects in the thematic selection task is domain-specific, renders their approach too inflexible to be a viable explanation of how we reason. Once again, this criticism is not disastrous. Much has been made of Fodor's distinction between impenetrable, and modularised, input systems and the high-level central systems (such as thinking) which he claims to be domain-general. His assumption that higher systems must be domain-general is based on a further assumption that all of our knowledge must be
available to us for a task such as thinking. However, there has been much progress since 1983 and several authors (e.g. Gigerenzer, 1995; Sperber, 1994) have pointed out that modularity does not imply cognitive impenetrability. That is, there may exist modules which are organised vertically so as to allow the passing of information from one to the other. There also exists some evidence in the literature on categorisation that there are domain specific differences in the ways that social and object knowledge is organised (e.g. Wattenmaker, 1995). Whilst it is true that cognitively impenetrable, domain-specific modules would lead to a system which was hopelessly rigid, this is not necessarily assumed by Cheng and Holyoak's approach.

A much more serious criticism of schema theory is the claim that it does not clearly specify how knowledge structures are separated into domains. Once again Oaksford and Chater make this point very tellingly

A second reason to suspect that domain-specific approaches are inadequate concerns the lack of any general principles concerning how an appropriate compartmentalisation of knowledge is to be achieved. (pg. 47)

This point has also been made by Johnson-Laird and Byrne (1991). The central issue here is that there are many ways in which knowledge may be structured and, without more evidence, there is no a priori reason to accept Cheng and Holyoak's version. All workers in the field agree that the way in which our knowledge of the world is divided up is crucial in determining how we reason and how we select information to reason with. However, several authors have suggested that pragmatic reasoning schemas are not the most parsimonious means of answering these questions.

At this stage in the review only preliminary conclusions may be drawn. What is very clear is the importance of the questions which Cheng and Holyoak have raised. As Gigerenzer (1995) astutely pointed out, human reasoning cannot be studied in isolation from "the content of the Ps and Qs" and the social context in which it takes place. The ways in which knowledge is both organised in, and retrieved from, long term memory are
crucial to any account of human reasoning. That the application of what we know about pragmatics will be instrumental in understanding these processes is unquestionable. What is questionable, however, is the extent to which Cheng and Holyoak have found the right approach. All workers in the area seem happy to admit the importance of the study of rules, obligations, and regulations in our understanding of human reasoning and evidence selection. As is normal in science, these same researchers cannot agree on the details. Undoubtedly, Cheng and Holyoak need firstly to specify their account more tightly, giving a stronger rationale for cutting up our knowledge of the world as they do. Secondly, there is a need for extensive testing of schema theory in settings which do not involve the Wason selection task.

### 1.3.3 The Relevance Theoretic Approach

An alternative account of subjects' behaviour on Wason's Selection Task has been offered by Sperber, Cara, and Girotto (1995). Like Cheng and Holyoak's (1985) work, this account has its roots in the principles of pragmatics. It will be argued that Sperber et al's approach, whilst inadequate on its own, contains valuable insights into information selection for hypothesis evaluation. Specifically, the importance of their Cognitive Principle of Relevance to a comprehensive account of information selection will be emphasised. Before describing their account in more detail, the notion of Relevance must be explained.

As Oaksford and Chater (1995) have pointed out, a theoretical account of the relevance of information to an individual is needed because current theories of reasoning are computationally intractable. They all assume that subjects only represent the most relevant or plausible information from which to draw inferences but none of them specify how that relevant information is retrieved from memory. The work of Sperber and Wilson and colleagues demonstrates how the most relevant information might be brought to mind. They call their approach Relevance Theory.
Relevance Theory defines the relevance of a piece of information to an individual using the following constraints:

- the greater the cognitive effect of processing a piece of information, the greater its relevance
- the greater the processing effort required for processing a piece of information, the less is its relevance

These constraints are fundamental, not only to Relevance Theory, but also to the experiments to be described in the next chapter. They will also be revisited in Chapter 5. Some brief discussion of them here is warranted therefore.

Sperber and Wilson (1986,1995) point out that any new piece of information is processed against the background of already existing information (i.e. existing beliefs and conjectures). They claim that if bringing this new information and the context together provides the information processor with some cognitive effect which could not have been achieved from the either the new information or the context alone, then this new information is relevant to the information processor in this context. Cognitive effects which lead to relevance might consist of the addition, abandonment, or revision of beliefs. Sperber and Wilson also claim that relevance is a matter of degree so that the greater the cognitive effects achieved by bringing a new piece of information together with a context, the greater the relevance of that piece of information. Likewise, the greater the cognitive effort involved in bringing this new piece of information together with a context, the less is the relevance of that piece of information to the information processor.

Sperber, Cara, and Girotto (1995) illustrate the interaction of these constraints with the following example. Suppose you are standing on a train platform and you want to catch the next train to Manchester. You don't have a train timetable however so, being told that the next train to Manchester leaves at 5.30 is relevant to you. This is because you are able to bring the already existing information that you want to go to Manchester as soon as is possible together with this new information to infer that the earliest that you can possibly
leave for Manchester is 5.30. Being told, on the other hand, that the next train to Manchester leaves after 4.00, whilst still relevant, is not as relevant. This is so if only because the proposition that the train leaves at 5.30 implies the proposition that the train leaves after 4.00, but not conversely. This part of the example concerns how cognitive effect might be a matter of degree.

The other part of their example concerns cognitive effort and goes as follows. Suppose that you are once again on the train platform (mentioned in the previous paragraph) and were in the same predicament - you want to go to Manchester but you don't have a timetable. Somebody tells you that the next train to Manchester leaves at 7500 seconds after 3.25. If you take the time, you can infer from this new piece of information that the next train leaves at 5.30, as well as all the information which follows from knowing that the next train leaves at 5.30. However, because there is extra processing effort involved in this case, it is not as relevant as the former piece of information.

All of the above leads to "the first (or Cognitive) Principle of Relevance" (Sperber, Cara, and Girotto, 1995) which states that "people tend to pay attention, at any given time, to the most relevant information available to them at the time, and to bring to bear on it a context of assumptions that will maximise its relevance" (pg. 49). The "Second (or Communicative) Principle of Relevance" (Sperber et al 1995) which was originally called the Principle of Relevance (Sperber and Wilson, 1986) involves differentiating communicated information from information present in the environment. It says that "every act of communication carries a presumption of its own relevance" (pg. 50). This is because a communicator, in order to facilitate the act of communication, requests the attention of her audience. She wants her audience to presume that what she requests their attention for is relevant.

This Communicative Principle of Relevance has both an effect, and an effort, component. These lead to the following "rational" comprehension strategy:
(1) the consideration of possible cognitive effects in the order of their accessibility
   (i.e. exerting as little cognitive effort as possible)
(2) stopping when the expected level of relevance is achieved

Following such a strategy is claimed to be rational because firstly, the addressee of any utterance may presume that the speaker has attempted to minimise the effort needed to arrive at the intended interpretation of their utterance and secondly, because the addressee need not assume maximal effect for the utterance. They need assume only adequate effect.

Given this brief outline of the basics of the relevance theoretic approach to communication, we may now turn to how it attempts to deal with the selection task. It does so in a very simple manner, by assuming firstly, that subjects understand the task as one of selecting relevant evidence for testing a rule and that this rule may only be tested indirectly (i.e. through its observationally testable consequences). Secondly, inferring some of the consequences of the rule is done spontaneously, as part of the process of comprehension. These spontaneous inferences lead to an interpretation of the rule which contains certain expectations of relevance (i.e. expectations, on the part of the subjects, about what is relevant to the experimenter). Also subjects trust both these relevance expectations and the processes which lead to them. Finally, subjects will select the cards which, if they could be observed, would lead to the testing of these spontaneously inferred consequences.

As may be seen from this account, the crucial factor is the spontaneous inferences which subjects make whilst comprehending the rule. This is what determines the way in which any conditional rule is found to be relevant. Sperber et al claim that these spontaneous inferences are predictable from the content of the rule and the context in which it is presented. To demonstrate this they concentrate on situations where subjects will interpret the task rule as a denial of the occurrence of $P$-and-$\neg Q$ cases. As they point out however, working from their effect and effort constraints, there are two reasons why subjects will generally fail to interpret the rule in this way. Firstly, it requires more processing effort to represent the rule as a denial of the occurrence of $P$-and-$\neg Q$ cases. This is because such an interpretation involves two negations, the implicit propositional
attitude of denial, and the explicit denial of $Q$. As Relevance Theory holds that processing effort is used as a source of information about the intended interpretation, then the extra effort involved in representing the rule in this way will suggest that this is not the intended interpretation of the rule.

Secondly, the normal effect of a conditional statement is the inference of $Q$ from a combination of the rule and the premise $P$ or the inference that there exist cases of $P$ and $Q$ from a combination of the rule and an assumption of relevance. This leads to two expectations. The first of these is compatible with the standard modus ponens inference:

If there is a $P$ then there is a $Q$
There is a $P$
Therefore, there is a $Q$

The second of these expectations corresponds to an invalid Denial of the Consequent inference:

If there is a $P$ then there is a $Q$
There is a $Q$
Therefore, there is a $P$

Accordingly, Sperber et al argue that in order to prevent subjects from interpreting the task rule in the usual way it is necessary to manipulate both the effect of the conditional involved and the effort involved in interpreting the conditional as denying the existence of cases of $P$-\textit{and-}(not-$Q$).

Sperber et al propose what they call "a recipe for constructing easy versions of the task" (Pg. 58). As regards cognitive effort, this involves selecting a pair of features $P$ and $Q$ such that the instance $P$-\textit{and-}(not-$Q$) is as easy, or easier, to represent than the instance $P$-\textit{and-}$Q$. As regards cognitive effect the recipe prescribes a context where knowing of the existence of $P$-\textit{and-}(not-$Q$) cases will lead to cognitive effects equal to, or greater than,
those derived from knowing of the existence of $P$-and-$Q$ cases. These authors present a series of four experiments along with their theoretical re-analysis of the selection task. Each of these experiments manipulates either the ease with which an instance of $P$-and-($not$-$Q$) may be represented or the effects to be derived from knowing of the existence of a $P$-and-($not$-$Q$) instance. In each case subjects' patterns of evidence selection correspond closely to those predicted by the authors. Four different effort manipulations were found to significantly facilitate subjects' selections. These were (1) including a $P$-and-($not$-$Q$) case in the text of the problem; (2) ensuring that the $P$-and-($not$-$Q$) case corresponded to a lexicalised item e.g. someone who is male ($P$) and unmarried ($not$-$Q$) is a bachelor; (3) giving subjects an intermediate task where they had to single out $P$-and-($not$-$Q$) instances; (4) creating a simple universe where $P$ and $Q$ are complemented by $R$ and $S$ which correspond to $not$-$P$ and $not$-$Q$ respectively. Sperber et al argue that this makes all four possible combinations of features equally easy to represent. On the side of cognitive effect two manipulations were found to significantly facilitate evidence selection. These were (1) making the existence of $P$-and-$Q$ cases trivial whilst the existence of $P$-and-($not$-$Q$) cases is unlikely and contentious; (2) making the existence of $P$-and-($not$-$Q$) cases diagnostic of something in context.

The relevance theoretic approach to the selection task has several advantages over the pragmatic reasoning schemas approach. Firstly, it can account for a lot more of the data than can pragmatic reasoning schemas. Secondly, the basic approach has far wider implications for cognitive psychology than does the notion of pragmatic reasoning schemas. The idea that subjects' behaviour on the selection task is due to a fundamental cognitive tension between effect and effort is a very attractive one indeed.

However, there is an apparent problem with relevance theory. It is possible that the approach is simply too broad to be of any real use. The real usefulness of relevance theory is that it specifies a constraint on the retrieval of background knowledge from long-term memory. This of course, is also true of the pragmatic reasoning schemas approach. As has already been seen however, a schema based approach is simply too rigid to be of much use. The problem with a relevance theoretic approach is that it merely specifies a general
principle and is unable to make more specific predictions about the characteristics of information which will be important in any given context. A much more specific alternative has been presented by Oaksford and Chater (1994). This will now be presented.

1.4. SUBJECTIVE PROBABILITY AND THE SELECTION TASK

1.4.1. Oaksford and Chater's Rational Analysis

Oaksford and Chater (1994) claim that the approach which they have adopted to the selection task accounts for not only the results on both abstract and thematic selection tasks discussed above but also for results on a variety of other selection task experiments which will be discussed below. Their approach will be critically evaluated in some detail here. This is because of its direct relevance to the experiments which will be presented in Chapter 2. Its direct relevance is due to the emphasis which it places on subjects' use of their background knowledge about the probabilities of events in their environment when selecting information with which to evaluate hypotheses. Oaksford and Chater's (1994) account follows the account given by Anderson (1990, 1991) of memory, categorisation, and problem solving. What these accounts have in common is their assumption that the attribution of rationality to an organism is dependent on whether that organism's behaviour is optimally adapted to its environment. Accordingly, Oaksford and Chater reject logic as a normative standard for subjects' behaviour on the Wason selection task and instead propose an alternative Bayesian account based upon the principles of optimal data selection (Federov, 1972; MacKay, 1992).

The basic principle behind optimal data selection is very simple and may be illustrated by the following example. Suppose you are asked to test the rule "If an individual is exposed to radiation then that individual will develop cancer". There are four classes of individuals one could examine to test the rule: individuals who have been exposed to radiation (p); individuals who have not been exposed to radiation (not-p); individuals who have developed cancer (q); individuals who have not developed cancer (not-q). The question, of course, is one of deciding which classes of individuals will supply
the most useful information. Intuitively, \( p, q, \) and \( \text{not-}q \) individuals are likely to be informative whereas not-\( p \) individuals will supply no information whatsoever. Oaksford and Chater have formalised these intuitions using Shannon-Wiener information theory and elements from the probability calculus. These intuitions are formalised differently for abstract and thematic versions of the Wason selection task and so Oaksford and Chater's accounts of these separate tasks will be dealt with separately. First however, the general approach will be further explained.

Oaksford and Chater claim that hypothesis testers should choose experiments to provide the greatest possible "expected information gain" when trying to decide between two hypotheses. In the case of the selection task these hypotheses are (1) that \( p \) is invariably associated with \( q \) (henceforth this hypothesis will be referred to as \( H_1 \)) and (2) that \( p \) and \( q \) are independent (this will be referred to as \( H_2 \)). For each of these hypotheses they derive a probability model from both the prior probability that each hypothesis is the case (which they assume to be equal in both cases i.e. = .5) and the probabilities of \( p \) and \( q \) in the task rule. Next, they define information gain as the difference between uncertainty before and after the selection of a card where uncertainty is defined using Shannon-Wiener information. Therefore, given \( n \) mutually exclusive and exhaustive hypotheses (\( H_i \)) the information gain which a piece of data (\( D \)) leads to is defined as:

\[
\text{Information before receiving } D: I(H_i) = -\sum_{i=1}^{n} P(H_i) \log_2 P(H_i)
\]

\[
\text{Information after receiving } D: I(H_i/D) = -\sum_{i=1}^{n} P(H_i/D) \log_2 P(H_i/D)
\]

\[
\text{Information gain: } I_g = I(H_i/D) - I(H_i)
\]

\( P(H_i/D) \) is calculated using Bayes theorem (for a discussion of Bayes theorem see a later section). Therefore, information gain may be thought of as the difference between the information contained in the prior probability of \( H_i \) and the information contained in the posterior probability of \( H_i \) given some piece of evidence \( D \).
However, when choosing which piece of evidence (card) to select, subjects do not know what the piece of evidence will actually be. Accordingly, Oaksford and Chater claim that subjects will calculate expected information gain with respect to all possible outcomes (i.e. for the p card the two possible outcomes are q and not-q). Oaksford and Chater's analysis leads to the following general predictions about the informativeness of the four cards on the selection task:

P Card: Is informative when P(q) is low. Its informativeness is largely independent of P(p)
Q Card: Is informative when P(p) and P(q) are both small
Not-q Card: Is informative when P(p) is low. Its informativeness is independent of P(q)
Not-p Card: This card is not informative.

In order to model the experimental data Oaksford and Chater are forced to make four assumptions. The first of these is that P(p) and P(q) are assumed by subjects to be low. They call this their "rarity assumption" and justify its application by reference to Klayman and Ha's (1987) "minority phenomenon" assumption and Anderson's (1990) work on causal inference. The justification for, and implications of, this assumption will now be discussed.

The rarity assumption is invoked only to model data from the abstract selection task. The argument is that most lexicalised properties refer to objects and events that are rare in our environment and on the abstract selection task subjects will use the default assumption that the tokens in the rule are rare. However, on the thematic selection task content and context will combine to provide information for the subject about the likely probability of p and q. Whether these are perceived to be rare will therefore, depend on the background knowledge which subjects bring to bear on the task.

There are several problems with this assumption. The first is the selective application of the rarity assumption. Two questions arise here. Firstly, is it not possible (as Kirby, 1994a suggests) that subjects may be using what they know about items in the abstract rule (i.e. Vowels, consonants, letters, even and odd numbers, etc.) to determine
their likely probability of occurrence? The second question which this selective application of the principle raises is why subjects can deal so easily with thematic rules if occasionally they do not respect rarity? And why bother to assume that some subjects will understand the items in a thematic rule to be common? If rarity is such an over-riding principle, surely subjects will assume these items to be rare. After all, they are generally not given enough information to be sure.

The second problem with the rarity assumption is its speculative nature. As Oaksford and Chater correctly point out, an environmental analysis of the type suggested by Anderson (1990) would be required to test the validity of the assumption. However, as Anderson points out, there are various types of converging analysis which may be carried out. Oaksford and Chater are unable to offer evidence from any sort of analysis. They cite Klayman and Ha's (1987) equally speculative assumption about the "minority phenomenon". Klayman and Ha do show that the structure of a rule discovery task, such as the Wason 2 4 6 task, changes if the probability of an event being governed by the rule goes over .5. It is arguable as to whether a probability of .5 constitutes rarity however. What is certain is that all of the probability values which Oaksford and Chater use in their analysis of the abstract selection task are far below .5.

The second assumption which Oaksford and Chater make is that P(q) ≥ P(p). They argue that this is a reasonable assumption because if P(q) ≤ P(p) then the hypothesis that q is invariably associated with p cannot hold. Once again, this is a problematic assumption. This is because it relies on a completely circular argument. The assumption concerns the subjects' perceived base rates for the items in the rule. If the base rate for q is perceived to be less than that for p then presumably, subjects would have more confidence in the hypothesis that p and q are independent (i.e. the task rule does not hold). However, subjects in most cases do not have any information about the base rates of the items in the rules typically used in the task. Consider the abstract rule: If there is an A on one side of the card then there is a 3 on the other side of the card. Subjects in most cases have information about neither the population from which the cards have been taken nor the base rates of certain attributes amongst that population. A similar (although weaker) argument may be made...
with respect to many of the thematic materials used in the literature. The point is that although subjects may realise that in order for the task rule to hold, \( P(q) \) must be greater than, or equal to, \( P(p) \), they do not know that this is the case. The assumption that subjects use a base rate of \( q \) which is greater than or equal to their base rate for \( p \) suggests that they are assuming something which they do not know to be true in order to discover what is most likely to be true. Any subject adopting this strategy would, at the end of her analysis, know the information gain to be expected from each card assuming \( P(q) \geq P(p) \).

The third assumption which Oaksford and Chater make is that every card has the possibility of being chosen. They make this assumption because although their initial model claims that the not-\( p \) card is never informative, nevertheless a small percentage of subjects do select that card. They assume that this is the case because not every subject will perform the appropriate, or any, analysis. To counteract this they add a constant (.1) to the \( E(I_g) \)s for each card. Although Oaksford and Chater are very likely to be correct in assuming that some subjects do not perform any analysis on the selection task, their method for allowing for this seems inappropriate. In effect, they are attempting to model the behaviour of two separate groups of subjects using the notion of expected information gain, whilst accepting that only one of these groups actually computes expected information gain.

This point has two repercussions for the theory. Firstly, it is likely that much more accurate predictions could be made using the Oaksford and Chater model if the card selections of successful subjects only was modelled. Secondly, because the model is not concerned with the mechanisms involved in deciding which cards to select, it has very little to say about erroneous (in its own terms) card selections. Although the notion of mechanism will be discussed in greater detail below, it is worth commenting upon briefly here. Evans and Over (1996a) have pointed out that the Oaksford and Chater model constitutes an alternative normative theory. Although they also point out that this alternative theory approximates more closely subjects' card selections than does Popper's (1959) notion of falsification (i.e. the old normative theory), nevertheless, Oaksford and Chater's account is primarily one of \textit{what} subjects should do rather than an account of \textit{how} they do it. This neglect of the mechanisms involved in selecting information to test
hypotheses means that the model very often cannot account for the behaviour of subjects who fail to behave as the model says they should.

The fourth assumption which Oaksford and Chater make is that "card choice is a competitive matter" (pg. 614). By this they mean that the less distinguishable a card is from alternatives the more likely it is to be selected. Before considering the steps which Oaksford and Chater take to implement this assumption in their model it should be pointed out that even taken at face value it is questionable. For example, it would be just as easy to argue that the more pronounced the difference in the expected information gain of two cards, the more likely it is that the card with the higher expectation gain would be chosen. Indeed there is evidence (Tversky and Shafir, 1992) that subjects find it very difficult to choose between two similar, or equally valued, options. Of course, if the selection task is properly viewed as a decision task, these competing predictions would be very easy to test. However, there is one very instructive mismatch between actual data and predicted expected information gain contained in Oaksford and Chater's paper. This mismatch exists in their attempts to model the data from 13 studies, reporting 34 standard abstract selection tasks, and involving 845 subjects. The predicted expected information gains for each card and the actual proportion of times each card was selected in these studies is given in Table 1.1.

<table>
<thead>
<tr>
<th></th>
<th>P Card</th>
<th>Q Card</th>
<th>Not-Q Card</th>
<th>Not-P Card</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted $E(I_g)$</td>
<td>.76</td>
<td>.20</td>
<td>.09</td>
<td>.0</td>
</tr>
<tr>
<td>Prop. selected</td>
<td>.89</td>
<td>.62</td>
<td>.25</td>
<td>.16</td>
</tr>
</tbody>
</table>

Table 1.1: Predicted expected information gain versus actual proportion of cards selected from Oaksford and Chater's (1994) meta-analysis of results on the standard abstract selection task.

The first noticeable feature of Table 1.1 is the contrast between the predicted $E(I_g)$ for the not-$p$ card and the actual proportion of all not-$p$ cards which were selected. This is because Oaksford and Chater did not implement their assumption about some subjects failing to analyse the task appropriately when modelling this data. It does however, make the point that subjects seem to fall into two separate groups on the task. More importantly however, is the disparity between the $E(I_g)$s for the $p$ and $q$ cards and the rate at which they
were selected. Although the $p$ card is predicted to have an expected information gain which is almost four times that of the $q$ card, the $q$ card is selected on almost two thirds of all occasions. It is obviously not the case therefore, that the more similar the $E(I_g)$ for a card is to the alternatives, the more often it is likely to be chosen. In fact, there is no obvious relationship between the $E(I_g)$s predicted by Oaksford and Chater and the rate at which subjects choose the different cards, other than the fact that the orderings are the same. Nevertheless, in order to model a different set of findings from the literature, Oaksford and Chater scale $E(I_g)$s by the mean information available. In other words, they divide the $E(I_g)$ for each card by the mean information gain for all of the cards. In this way they reduce the differences in $E(I_g)$ between all of the cards. This is worrying for two reasons. Firstly, it does not seem to be warranted by the data. Secondly, it is a procedure which is used when modelling some aspects of the literature and not others. This second problem directly leads to a third problem. Because of their seemingly indiscriminate use of this scaling procedure, Oaksford and Chater seem to be suggesting that more processing takes place on some tasks than on others. There are absolutely no a priori grounds for such a suggestion.

Having examined in detail the assumptions underlying the Oaksford and Chater model of selection task performance it is now time to consider how well it actually deals with the data. Obviously, an attempt to discuss all of the data is out of the question here. It is worth mentioning however, the variety of selection task variants for which Oaksford and Chater have attempted to account. As discussed above, they have modelled results on the standard abstract selection task. They have also modelled the finding that card selections on the abstract task are non-independent (Pollard, 1985) and what they claim to be the findings on the negations paradigm (Evans and Lynch, 1973; Manktelow and Evans, 1979; Evans, 1989). In the negations paradigm the antecedent and the consequent of a rule can contain negated constituents. This leads to four possible selection task rules (if $p$ then $q$; if not-$p$ then $q$; if $p$ then not-$q$; if not-$p$ then not-$q$). Initially it was found (Evans and Lynch, 1973) that subjects selected those cards which matched the terms in the rule. Later work (Manktelow and Evans, 1979) suggested that matching mainly occurs for the consequent card whereas antecedent card selections tend to be in accord with logical case. The order of predicted $E(I_g)$ for the cards in this paradigm agree with the rates at which cards tend to be
selected. Likewise, the Oaksford and Chater model also predicts the order in which cards
tend to be selected in the various studies (e.g. Reich and Ruth, 1982; Oaksford and
Stenning, 1992) which have managed to suppress this matching effect. However, it should
be noted that there is some uncertainty as to whether matching mainly occurs for the
consequent card only. Based on a meta-analysis of four experiments, Evans, Newstead and
Byrne (1993) concluded that matching occurs for all four logical cases. On the other hand,
the data upon which Oaksford and Chater base their meta-analysis does not include one of
the studies considered by Evans, Newstead and Byrne, but does include a further three
experiments, two of which are contained in Oaksford and Stenning (1992) and one of which
is an unpublished control experiment.

In order to model results on the thematic selection task, Oaksford and Chater have
introduced the notion of subjective expected utility (for a further discussion of subjective
expected utility, see the General Discussion of Chapter 2) and argue that subjects on
thematic variations of the task are attempting to maximise expected utility. The basic
notion is the utility of both the condition and the action to those individuals governed by the
rule. To model thematic facilitation Oaksford and Chater weight the utility of a piece of
information by the probability that checking the relevant card will produce that information.
This leads to the following four equations which express the expected utilities of each card
from the enforcer's and the actor's perspective (where "con" refers to condition, and "act"
refers to action):

\[
EU(\text{con}) = P(\text{act} | \text{con})U(\text{con}, \text{act}) + P(\text{act} | \text{con})U(\text{con}, \text{act}) \quad \text{eq. 1}
\]

\[
EU(\overline{\text{con}}) = P(\text{act} | \overline{\text{con}})U(\overline{\text{con}}, \text{act}) + P(\text{act} | \overline{\text{con}})U(\overline{\text{con}}, \text{act}) \quad \text{eq. 2}
\]

\[
EU(\text{act}) = P(\text{con} | \text{act})U(\text{con}, \text{act}) + P(\text{con} | \text{act})U(\text{con}, \text{act}) \quad \text{eq. 3}
\]

\[
EU(\overline{\text{act}}) = P(\text{con} | \overline{\text{act}})U(\text{con}, \text{act}) + P(\text{con} | \overline{\text{act}})U(\text{con}, \text{act}) \quad \text{eq. 4}
\]

This part of the Oaksford and Chater account is relatively uncontroversial. The majority
consensus in the field at the moment is that expected utility is crucial in understanding the

Despite Oaksford and Chater's uncontroversial use of the notion of expected utility in modelling subjects' performance on thematic versions of the selection task, there still remains a core problem with attempting to compare the Oaksford and Chater model with any other in the reasoning literature. The problem is one of incommensurability (see Brown, 1988 for a discussion of this issue). Essentially, Oaksford and Cater disagree with the majority of workers in the area as to what constitutes the proper subject matter for the scientific study of evidence selection. That this is true is borne out by reference to Oaksford and Chater's claim that "The purpose of a rational analysis is to show that behaviour is optimally adapted to the environment" (pg. 609). Further evidence of the incommensurability of the positions is provided by Anderson's (1990) claim that a rational analysis constitutes a theory at the "computational level" (see Marr, 1982) and helps define the issues in developing a mechanistic theory of the behaviour being studied. It is inappropriate to argue therefore (as both Oaksford and Chater and Holyoak and Cheng have done) that their theories are interchangeable. They are simply theories at a different level. One seeks to justify observed patterns of behaviour by recourse to an analysis of the environment (in this context Evans and Over's characterisation of Oaksford and Chater's account as a normative one makes eminent sense) whilst the other seeks to present a mechanistic account of how that behaviour is produced.

Unfortunately for Oaksford and Chater however, their analysis does not confine itself to the environment. At each step they have been forced to make assumptions about psychological processes. All of the criticisms which have been made above of the assumptions underlying their analysis are concerned with this point. At worst, as with the assumption that card selection is competitive, they make very questionable assumptions about the cognitive factors underlying behaviour, and at best, as with the notion of "scaled" $E(I_g)$, they implement these assumptions without regard for processing constraints.
That there is such confusion within the theory argues against its acceptance. Some workers in the field have also argued that the model is unsuccessful in accounting for data which exists in the literature. For example, Evans and Over (1996a&b) have argued that the model does not account for results found by Pollard and Evans (1983) and Kirby (1994) (see the next section for a discussion of these studies). Almor and Sloman (1996) complain that the theory cannot account for their data. Laming (1996), whilst making some of the same criticisms made here, emphasises the fact that

*Oaksford and Chater propose a theory with three free parameters, to which they assign particular values, and other assumptions besides, to achieve merely rank order agreement with four observed frequencies* (pg. 365).

He argues that a rank order correspondence may be achieved in many ways and, on its own, does not justify the theoretical conclusions which Oaksford and Chater draw.

Although the debate is ongoing about whether Oaksford and Chater's model does, or indeed needs to (Oaksford and Chater, 1996), account for this data, it will not be described in detail here. Instead the work of Kirby (1994a) and Pollard and Evans (1983) will be discussed in detail in the next section. Suffice it to say that Oaksford and Chater's work has been very controversial and there is by no means agreement amongst workers in the field about its status. Oaksford and Chater themselves are not clear as to the level at which they want their account to work.

**1.4.3 The effect of subjective probability**

The central tenet of Oaksford and Chater's thesis about non-deontic selection tasks is that subjects are sensitive to the probabilities of events occurring in their environment. There exist several sources of evidence which suggest that they are correct in this assumption. For example, in the concept literature there is a long history of research into the importance of feature frequency. Several authors have pointed out that since frequent features occur consistently amongst positive instances of a concept whilst infrequent
features occur inconsistently, the defining features of a concept may be identified by accumulating frequency information (e.g. Bourne and Restle, 1959; Hull, 1920; Restle, 1955). There has been some recent work which supports this view. Kellog (1980) has demonstrated the importance of high frequency features on an abstract concept learning task, whilst Bourne (1982) and Neumann (1974, 1977) have shown that single feature frequencies are encoded in concept learning tasks. Even more recently, Wattenmaker (1993) found that single feature frequencies were encoded in both intentional and incidental concept learning tasks.

A second line of research which suggests that subjects are sensitive to the probabilities of events occurring in their environment comes from the extensive work done on implicit learning (see Reber, 1994 and Berry and Diennes, 1993 for reviews and discussion). Reber and Millward's work using a probability learning paradigm is especially instructive in this regard. Reber and Millward (1971) found that subjects can accurately anticipate the changing probabilities of events even when the anticipatory response requires an integration of information across 50 preceding events. Millward and Reber (1972) reported results which suggested that subjects were sensitive to dependencies between events where there were up to six intervening events between the dependent events.

The importance of these demonstrations of the effects of feature frequency on subjects' acquisition of concepts and prediction of future events to the current discussion should be clear. The more frequently an event or feature occurs the more probable it may be said to be. The two lines of research briefly described above suggest that subjects are sensitive to the probabilities of events in their environment. If this is the case, then it would be expected that the probabilities of the events, or items, which are governed by the hypotheses which subjects are reasoning about, should have an effect on the information selected to test those hypotheses. Despite the centrality of this assumption to Oaksford and Chater's model, surprisingly little research has been done on this question. As the notion of subjective probability will be crucial to an understanding of the experiments to be described in Chapter 2, the work which has been done will now be reviewed.
The first study which will be discussed is that of Kirby (1994a). In the experiments contained in this study he investigated the effect of subjective probability on subjects' pattern of card selection on the indicative selection task. All of these experiments were based upon a re-analysis of Wason's (1966) original selection task. The original rule which Wason used was "If a card has a vowel on one side then it has an even number on the other side". He points out that the set of $P$-instances (all 5 vowels) is smaller than the set of $not-P$ instances (the 21 consonants). Therefore, the fact that Wason (1968) reported only a minority of subjects choosing the $not-Q$ card may have been due to the fact that the posterior probability (the prior odds) of finding a $P$ on the back of the not-$Q$ card were low. Accordingly, Kirby reports the results of three experiments where the relative sizes of the $not-P$ and the $P$ set were varied.

In the first of these experiments subjects were given a scenario concerning a computer which had been given the task of printing cards with an integer from 0 to 1000 on one side and one of two arithmetic symbols (+ or -) on the other side. The rule for the small $P$ set condition was: "If the card has a 0 on one side, then it has a + on the other side". The rule for the large $P$ set condition was: "If the card has a number from 1 to 1000 on one side, then it has a + on the other side". Thus, the likelihood of finding a $P$ on the back of the not-$Q$ card was higher in the large $P$ set than it was in the small $P$ set condition. If subjective probability does affect subjects' patterns of evidence selection then choice of the not-$Q$ card should be higher in the large $P$ set condition than in the small $P$ set condition. This is exactly what was found. A total of 73% of subjects ($N = 44$) selected the not-$Q$ card in the large $P$ set condition versus 49% ($N = 45$) in the small $P$ set condition. This result was replicated in two subsequent experiments. As the size of the $P$ set increases, so too does the proportion of subjects selecting the not-$Q$ card.

Unfortunately however, there is a problem with the design of Kirby's experiments (Over and Evans, 1994; Kirby, 1994b). Over and Evans argue that there is a plausible alternative reading of Kirby's task which would lead subjects in the small $P$ set to conclude that there is only one instance of $P$ amongst the cards. Remember that subjects were told
that the computer was given the task of printing out cards with an integer from 0 to 1000 on one side. Thus subjects may have inferred that there were 1001 cards in total.

Remember also that the rule concerned cards with a 0 (P) on one side and a + on the other side (Q). As the P-card is always one of the cards from which selections are allowed, subjects may have inferred that there could not be a P on the back of the not-Q card because they could already see the front of the card with a P. Such a reading of the task would lead subjects not to choose the not-Q card which confounds the set size explanation offered by Kirby.

Fortunately, Over and Evans are able to cite two other studies which support Kirby's hypothesis. The first of these is Evans and Pollard (1981). These authors presented subjects with indicative conditional rules the truth or falsity of which had been independently rated. They found that subjects were more likely to select the not-Q card when the conditional rule being tested was believed to be false. Evans and Pollard interpreted this finding as being due to the falsifying case of P and not-Q being readily available to subjects who did believe the conditional rule to be false. As Over and Evans (1994) point out, this explanation is essentially the same as that offered by Kirby for his results as it makes the selection of the not-Q card dependent upon subjects' expectations of finding a P.

The second study which Over and Evans cite is that of Pollard and Evans (1983). This study also demonstrated facilitation of false consequent selections but this time a probability learning paradigm was employed. The rules used were abstract with and without negative components, and the probabilities of the combinations of values on the cards were established prior to the selection task. The materials which Pollard and Evans used consisted of sets of cards with red and blue sides. For example, subjects could be shown a card with a red side which had a triangle or was blank and a blue side which had a star or was blank. Selection task conditionals such as:

If there is a triangle on one side then there is a star on the other side
were established with different contingencies in the probability learning task. For instance, the *usually true* conditional was composed of

- seven cards with a $P$ on one side and a $Q$ on the other
- one card with a $P$ on one side and a *not-Q* on the other
- seven cards with a *not-P* on one side and a $Q$ on the other
- one card with a *not-P* on one side and *not-Q* on the other

In the above case the probability of both *not-Q* given $P$, and $P$ given *not-Q*, was $1/8$. For the *usually false* conditionals on the other hand, both of the equivalent probabilities were $7/8$. As predicted, significantly more *not-Q* card selections were made with *usually false* conditionals.

As previously mentioned, there is a problem with the design of Kirby's experiments. Therefore, as Evans and Over (1996a) argue, it cannot be used as evidence for or against Oaksford and Chater's rational analysis. The Pollard and Evans study on the other hand, is sound. Unfortunately however, there is very little agreement about how it should be modelled. Oaksford and Chater claim that Pollard and Evans' results are predicted by their model whilst Evans and Over (1996a and b) argue that Oaksford and Chater only succeed in modelling the data by making highly questionable adjustments to their parameter values.

It would seem then that there is no existing empirical evidence, the status of which all concerned parties agree on, which may be used to test the Oaksford and Chater model. However, the point of this section is not to arbitrate between Oaksford and Chater's account and its competitors. Rather it is to establish that subjective probability does have an effect on subjects' patterns of card selections. In the light of the evidence discussed it would seem that subjective probability is an important factor in determining the rate at which subjects select the *not-Q* card. One of the aims of the first study to be described in the next chapter will be to demonstrate the importance of subjective probability in an
alternative selection task. It is to a description of this alternative selection task which we will now turn.

1.5 AN ALTERNATIVE SELECTION TASK

1.5.1 The pseudodiagnosticity paradigm: An alternative selection task

One of the few tasks, other than Wason's selection task, which has been used to investigate the evidence which subjects select to test hypotheses is the pseudodiagnosticity task (Doherty, Mynatt, Tweney and Schiavo, 1979). Pseudodiagnosticity is the descriptive term for subjects’ failure, when asked to select evidence needed to make a judgement under uncertainty, to select diagnostically relevant information. For example, subjects in the original Doherty et al study were asked to decide whether a pot recovered by a diver came from one of two islands - Coral Island or Shell Island. They were told that the pot possessed a list of eight characteristics such as being made from smooth clay and having curved handles. Subjects were then told that they could select information about the percentages of pots from each of the two islands which possessed these characteristics. Consider Bayes theorem, which states that

\[
\frac{P(H_1/D_i)}{P(H_2/D_i)} = \frac{P(D_i/H_1)}{P(D_i/H_2)} \times \frac{P(H_1)}{P(H_2)}
\]

where \(H\) and \(D\) stand for hypotheses and data respectively, the subscripts 1 and 2 label two exclusive and exhaustive hypotheses, and \(i\) labels a piece of evidence. \(P(H_1/D_i)\) is the posterior probability that \(H_1\) is the case, \(P(H_1)\) and \(P(H_2)\) are the base rates or prior probabilities of \(H_1\) and \(H_2\) respectively. \(P(D_i/H_1)\) and \(P(D_i/H_2)\) are the probabilities of the evidence given the hypotheses under test. In the example given above, the percentages of pots from Coral and Shell Islands which possess a given feature are equivalent to \(P(D_i/H_1)\) and \(P(D_i/H_2)\).

From the preceding discussion it is clear that subjects should select information about both the percentage of pots on Shell Island and on Coral Island which possess any
particular feature. Typically, this is not what subjects do. Instead, when allowed to choose six pieces of information, only 13% of Doherty et al's subjects chose evidence about three features relevant to both hypotheses. The vast majority of subjects did not choose any evidential pairs i.e. evidence about the percentage of pots on both islands possessing a particular feature.

This study by Doherty et al investigated subjects' choice of evidence to allow them to make a judgement under uncertainty. Fischhoff and Beyth-Marom (1983) used essentially the same task structure to examine subjects' use of normative information in making such a judgement. They found that subjects given a diagnostic likelihood ratio (i.e. \( P(D_i/H_1) > P(D_i/H_2) \) or \( P(D_i/H_1) < P(D_i/H_2) \)) responded in the direction appropriate to the relative likelihoods. Almost 50% of subjects who were given non diagnostic information \( (P(D_i/H_1) = P(D_i/H_2)) \) did not change their levels of belief in the focal hypothesis upon receipt of that information. Thus it would seem that subjects, when given the appropriate evidence, do have some working understanding of diagnosticity. This conclusion must be moderated in the light of some work by Ofir (1988) who found that the false alarm \( P(D_i/H_2) \) was only utilised in cases where there was inconsistency between the base rate and the likelihood for \( H_1 \). It seems that subjects will only use diagnostic information to resolve inconsistency and so do not possess an understanding of the diagnosticity of evidence.

1.5.2. The Pseudodiagnosticity task: A closer inspection

From the research described in the previous section two conclusions about the pseudodiagnosticity task may be drawn. Firstly, and most obviously, it is a selection task. Secondly, the uses to which it has been put are inextricably tied up with research into subjects' understanding of Bayes theorem. However, before it is put to use in investigating the factors underlying the selection of evidence for the evaluation of hypotheses, it must be re-analysed. This is because previous analyses of the task have been concerned with only one normative application of Bayes theorem. As Oaksford and Chater have suggested, subjects may be conforming to Bayes theorem in a non-obvious way.
The most simple version of the task is the one used by Mynatt, Doherty, and Dragan (1993). They presented subjects with a scenario which described an object (O) as belonging to one of two categories - X or Y (the hypothesis that O belongs to X will be referred to as $H_X$ whilst its complementary hypothesis will be referred to as $H_Y$) - and asked them to select information to help them decide between the hypotheses. Below is an example of such a scenario:

Your sister has a car she bought a couple of years ago. It's either a car X or a car Y but you can't remember which. You do remember that her car does over 25 miles per gallon and has not had any major mechanical problems in the two years she's owned it.

Next, subjects were told what percentage of car Xs do over 25 miles per gallon and were asked to choose one piece of information from the remaining three. Figure 1.2 illustrates how Mynatt et al conceptualised the task.

As may be seen from the figure below two of the four possible pieces of information are relevant to $H_X$ whilst the remaining two are relevant to $H_Y$. Mynatt, Doherty and Dragan (1993) found that, in a standard version of their task, the majority of subjects selected pseudodiagnostic information. Mynatt et al suggested focus as the cause of such pseudodiagnostic evidence selection. They claimed that underlying focus is subjects' inability to hold in working memory more than one hypothesis at a time. However, they also demonstrated that when evidence emerges which causes focus to shift from one alternative to its complement subjects choose diagnostic evidence. This demonstration consisted of giving subjects information about X which suggested that its complement was actually the case (e.g. by telling subjects that 35% of instances of car Xs do over 25 m.p.g.). They further demonstrated that focus may be manipulated onto an evidential feature rather than onto an alternative. For example, if the task of choosing between alternatives is given some utility for subjects (in this case they are told to imagine that they are going to act on their decision), then when given one piece of information
subjects will choose the information that allows them to make a comparison between the alternatives on that feature (i.e. they will choose the diagnostic evidence).

\[
\begin{array}{ c | c | c | c }
\hline
\text{Information} & \text{Car X (Hx)} & \text{Car Y (Hy)} \\
\hline
\% \text{ over 25 mpg} \quad \text{(D1)} & \text{Cell A} & \text{Cell B} \\
& p(D1/Hx) & p(D1/Hy) \\
\% \text{ no problems in two years} \quad \text{(D2)} & \text{Cell C} & \text{Cell D} \\
& p(D2/Hx) & p(D2/Hy) \\
\hline
\end{array}
\]

**Figure 1.2: Mynatt et al's (1993) conceptualisation of the pseudodiagnosticity task**

It is clear (see Evans and Over, 1996b for a discussion of this point) that Mynatt et al's interpretation of the pseudodiagnosticity task accords with a relevance account of information selection. In the standard case subjects choose an additional piece of information about the hypothesis which they are currently considering as it is the most relevant. However, when the decision between the two alternatives has some utility for subjects, they select the piece of information which will allow them to compare the alternatives. A similar interpretation of Mynatt et al's results has been offered by Evans and Over (1996b).

The problem with such an interpretation is that it fails to consider the probabilistic structure of the task. For instance, it may be the case that subjects will in all cases select the information which is most likely to discriminate between the hypotheses i.e. they are attempting to maximise expected information gain. Such a reading of the task would lead one to conclude that in the standard version of the task, subjects might tend to choose an
additional piece of information about the hypothesis they are currently considering because they perceive it to be less likely that instances of the first category will possess both features than it is that members of either category will possess just one of the features. This is expressed by Inequality 1:

\[ p(D_1 \& D_2) \leq p(D_1) \]  

Inequality 1 expresses the law of conjunction. The law of conjunction simply states that it is always less, or equally, likely that an object (or set of objects) will possess two features than it is that the object will possess just one of those features. As applied to the pseudodiagnosticity paradigm, this means that choosing \( p(D_2/H_x) \) might, according to an Oaksford and Chater type analysis, be perceived by subjects to be the evidence selection with the highest associated information gain. In other words, subjects may consider it to be less likely that a high percentage of instances of category X will possess both features than it is that a high percentage of instances of category Y will possess the first feature. Accordingly, a selection of information concerning \( p(D_2/H_x) \) may, in certain cases, be the one most likely to yield information which will discriminate between the hypotheses.

Although a thorough analysis, of the type suggested, will not be attempted here, it is clear that the probabilistic nature of the pseudodiagnosticity task has never received the attention it deserves. The fact that it has not received this attention suggests that Mynatt Doherty and Dragan's (1993) results should be interpreted with caution.

However, an analysis of the type suggested above rests on the assumption that subjects are aware of Inequality 1. There is much evidence that this is not the case. For example, Tversky and Kahneman (1983) have found that the majority of subjects violate the law of conjunction captured by Inequality 1. They gave subjects the following description:

*Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of*
discrimination and social justice, and also participated in anti-nuclear demonstrations.

Subjects were then asked to rank a series of statements in terms of their probability. Amongst the list were the following statements:

(A) Linda is a bank teller

(B) Linda is a bank teller and is active in the feminist movement

Subjects rated statement B as being more probable than statement A. This result held up even when a between subjects design was used where subjects had to rate a list of statements where either A or B had been deleted. Equivalent results have been found by Agnoli (1991) using a similar paradigm with children.

In the face of such evidence about the conjunction fallacy, it is unlikely that subjects typically choose evidence about $p(D_2/H_x)$ because they perceive it to be the selection most probable to be informative, although, for other reasons, they may perceive it as being the most relevant (see the next section for a discussion of Evans and Over, 1996b). However, the research described in the opening section of this chapter suggests that subjects are sensitive to the base rates of events occurring in the environment. If this is the case such a sensitivity should manifest itself on the pseudodiagnosticity task. The experiments to be described in the next chapter will examine the effect of subjective probability on subjects' information selections on the pseudodiagnosticity task.

1.6 SUMMARY AND CONCLUSIONS

This chapter has reviewed, in detail, three approaches to accounting for the effect of background knowledge on the selection of information for the evaluation of hypotheses. From the preceding review it is clear that the first of these approaches - Cheng and Holyoak's pragmatic reasoning schemas - was of central importance for its concentration
on pragmatics. Nevertheless, it does seem that the approach is too rigid to be ultimately successful. The level of organisation in background knowledge which it calls for, would, in the final analysis, seem too great. In addition, there is the fact that Cheng and Holyoak's account does not generalise past deontic selection tasks. Their account has to be viewed as being unparsimonious.

Much more promising however, are the approaches of Sperber and colleagues, and of Oaksford and Chater. Such approaches require much less rigidity from background knowledge. Both seem capable of accounting for the vast range of human competence in applying background knowledge in order to select information. As was discussed however, there are potentially serious problems with both approaches. The theory of Sperber et al fails to deliver anything concrete in terms of what characteristics of a piece of information will be important in any given context. It is undeniable that their work on the effect/effort trade-off is crucial in explaining why some information will be more important than other information in a given situation. Nevertheless, it does seem important to give an account (which will possibly have to be domain-specific) of the sources of information which are important to the organism. In a sense, this is precisely what Cheng and Holyoak have done. Unfortunately, they have committed the sin of over-specificity rather than over-generality.

The Oaksford and Chater approach, on the other hand, would seem, on the face of things, to have got the balance exactly right. They claim (Oaksford and Chater, 1995) to have produced a principled account of relevance on Wason's Selection Task. Certainly their concentration on a specific, but at the same time, general source of information for the subject (i.e. knowledge about subjective probability) seems laudable. As has been pointed out however, there exist serious problems with the specifics of their model. They are profligate with processing resources, introducing unprincipled extra stages of processing to model the results from individual tasks. Their rarity assumption, upon which so much of the model rests, seems questionable. Much more worrying however, is their apparent confusion about the level at which they want their account to work. Is it a normative account, as Evans and Over (1996a) have suggested? If so, then it is possible for their model to both preserve its normative status and work at a computational level (Marr,
1982). Unfortunately, Oaksford and Chater, as has previously been mentioned, do not confine their theorising to such a level of analysis. In order to make their equations work, they have had to make numerous assumptions which have consequences for the type, and amount, of processing required (a point also made by Laming, 1996). It would seem that, in order to explain all of the existing phenomena in terms of subjective probability, Oaksford and Chater have had to sacrifice plausibility.

Nowhere is it more apparent that subjective probability will ultimately prove to be less than the whole story than in the earlier consideration of the pseudodiagnosticity task. Although it was argued (and will be tested in Chapter 2) that subjective probability would prove to have an effect on subjects’ pattern of evidence selections on that task, it is obvious from the analysis contained in the previous section that subjective probability cannot be the whole story (in order for subjects to be selecting further information based on a probabilistic analysis they would have to be conforming to the law of conjunction which previous research makes unlikely). Even if subjective probability does turn out to affect subjects’ behaviour on the task, the basic finding of pseudodiagnosticity remains to be explained. It is likely that the explanations of Mynatt et al (1993) and Evans and Over (1996b) will turn out to be correct. Subjects choose further information about the hypothesis which they are currently considering for reasons of cognitive economy. Having constructed an initial representation involving that hypothesis, it requires less cognitive effort to select further information relevant to that representation than it does to consider the alternative hypothesis. Interestingly, Evans and Over (1996a&b) also recognise the profound importance of subjective probability to information selection. Accordingly, it is worth discussing in some more detail, the specifics of Evans and Over’s (1996b) account of how information from background knowledge informs the selection of information for hypothesis evaluation.

Evans and Over take as their starting point the fact that "all human thought is highly subject....to relevance effects" (pg. 45). By this they mean that humans reason about only a highly selective representation of both the information contained in the situation at hand and background knowledge. It is obvious that this starting point has much in common
with the three perspectives previously discussed in this chapter. However, it is in their
definition of what determines relevance that Evans and Over suggest a means of resolving
the tension between accounts which are too general on the one hand, and accounts which
are too specific on the other. Their account of the determinants of relevance has a decided
advantage over the accounts of both Sperber and colleagues and Oaksford and Chater. Its
advantage lies in the fact that it contains the best of both of the other accounts, and more.
For example, Evans and Over claim that there are linguistic determinants of relevance. It
will be remembered that these linguistic sources of relevance were discussed at the start of
this chapter in the context of Evans' (1984;1989) work on the abstract selection task.

In addition to linguistically-determined relevance, Evans and Over claim that
processing effort and constraints may also be a determinant of relevance. Once again, this
claim has already been discussed in relation to the pseudodiagnosticity paradigm. Along
with Mynatt, Doherty and Dragan (1993), Evans and Over claim that when selecting
information to enable them to decide between the alternatives present in the
pseudodiagnosticity task, subjects will select information relevant to the hypothesis which
they are currently considering. Evans and Over argue that this is due to a fundamental
constraint of the human information processing system: namely that, very often, subjects
can only attend to one hypothesis at a time.

As mentioned above, Evans and Over also emphasise the importance of subjective
probability to the selection of information for hypothesis evaluation. However, instead of
Oaksford and Chater's (1994) measure of Expected Information Gain, they suggest that the
concept of Epistemic Utility be used. In this context, epistemic utility may be defined as
the degree to which a piece of information is useful in deciding between two hypotheses,
and is specified in terms of the posterior probabilities of the hypotheses in the light of the
information. Although the precise measure which Evans and Over use to capture the
notion of epistemic utility will not be described here, it will be briefly discussed at the end
of Chapter 2. Most important in this context is not the measure which Evans and Over use,
but the fact that they do not argue that it fully captures our concept of epistemic utility
which they claim to be much richer than any fixed measure. Neither do they claim that
relevance is determined solely by our knowledge of the probabilities of the events in our environment. Instead, whilst recognising that some kind of principled account has to be given of the importance to the organism of specific sources of information (such as linguistic knowledge or background knowledge about subjective probabilities), Evans and Over also acknowledge the importance of the more general constraints of cognitive effort and cognitive effect.

Whilst it may seem strange to discuss an alternative theoretical account in the concluding section of this chapter, it does seem fitting as, by so doing, a conclusion becomes possible. This conclusion is that in order to fully understand the role played by background knowledge in subjects' selection of information for the evaluation of hypotheses, it is necessary to combine approaches which already exist in the literature. It is necessary to temper specific work on the role of subjective probability and information gain, with the general insight that cognitive effect must be balanced against cognitive effort. The theoretical work of Evans and Over suggests that this is possible. The first two experiments in this thesis will best be understood as an attempt to manipulate this balance.
CHAPTER 2 - EXPERIMENTS 1 AND 2: SUBJECTIVE PROBABILITY AND THE SELECTION OF EVIDENCE FOR THE EVALUATION OF HYPOTHESES

2.1 INTRODUCTION

In the previous chapter, several recent approaches to the question of how background knowledge affects subjects' selection of information for the evaluation of hypotheses were discussed. It was concluded that an account incorporating general relevance theoretic principles and specific assumptions concerning subjective probability (amongst other factors) was necessary. It was pointed out that Evans and Over (1996b) have demonstrated that such an account is possible. It was also claimed that the pseudodiagnosticity task was an ideal instrument for testing the necessity of such an account. The experiments to be reported in this chapter will constitute such a test. Before proceeding with a description of these experiments, some initial discussion is warranted.

As mentioned in Chapter 1, the version of the pseudodiagnosticity account used by Mynatt, Doherty and Dragan (1993) involved presenting subjects with scenarios such as the following:

*Your sister has a car she bought a couple of years ago. It's either a car X or a car Y but you can't remember which. You do remember that her car does over 25 miles per gallon and has not had any major mechanical problems in the two years she's owned it.*

Next, subjects are told the likelihood that members of one of the hypothesised categories possess one of the features also possessed by the target instance. They are then asked to select a further piece of information to help them decide between the hypothesised categories.
According to Bayes theorem, this additional piece of information should always enable subjects to compare the hypothesised categories across a single feature. Thus, if the initial piece of information concerned the percentage of model Xs which do 25 miles per gallon, then the subject should select an additional piece of information concerning the percentage of model Ys which do over 25 miles per gallon. Typically however, this is not the selection which subjects make. The majority of subjects select a further piece of information relevant to the hypothesised category about which they already possess some information.

Explaining this phenomenon is, as discussed in Chapter 1, very difficult for any account based solely on subjects' knowledge of subjective probability. As was pointed out, the explanation given by Mynatt et al for their results is very similar to that given by Evans and Over (1996b). Both sets of authors claim that subjects make the pseudodiagnostic selection for reasons of cognitive economy. Mynatt et al argued that subjects can hold in working memory, and operate upon, only one hypothesis at a time. In addition to this, they claimed that subjects have a bias to test the hypothesis they think to be true. Taken together, these assumptions account for the pseudodiagnostic selections observed on the task.

The Evans and Over argument is slightly different. These authors claim that the initial piece of information leads to the foregrounding of the hypothesis to which it relates. This leads directly to the typical selection of further information concerning the foregrounded hypothesis. It will be remembered that Mynatt et al also presented some results which showed that if the context in which the problem was presented assigned utility to the decision between the alternatives, then subjects were more likely to select the diagnostic information. Specifically, they told subjects to imagine not that the car was their sister's, but that they were thinking of buying the car. Once again, Evans and Over's explanation of these results is slightly different to that of Mynatt and colleagues. They argue that the wording of the problem indicates concern with the car's attributes. This leads to the foregrounding of the attribute which the initial piece of information concerns which, in turn, leads to the selection of the diagnostic evidence.
These explanations are similar in that they both place a premium on the notion of cognitive economy. The Evans and Over explanation seems to be more plausible however, as it suggests that the observed patterns of information selection are due to the manner in which subjects initially represent the problem. Although Mynatt et al's explanation does explicitly discuss the notion of cognitive effect (i.e. the utility of a comparison), Evans and Over successfully account for the results in terms of cognitive effort. However, the balance between cognitive effort and cognitive effect may very easily be investigated using this task.

Consider the subject presented with the following scenario:

*Your sister has a car she bought a couple of years ago. It's either a car X or a car Y but you can't remember which. You do remember that her car does over 100 miles per gallon and has not had any major mechanical problems in the two years she's owned it.*

Next the subject is told the percentage of model Xs which do over 100 miles per gallon. As usual, her task is to select a piece of information to help her decide between the hypotheses. Which piece of information should she select? For reasons of cognitive effort, she may select a further piece of information pertinent to model X cars. In this case however, she is more likely to accrue significant cognitive effects from the diagnostic selection. This is because from her background knowledge about cars, she knows that doing 100 miles to the gallon is, statistically, a very rare feature amongst the general population of cars. It is unlikely that very many model Y cars will do 100 miles per gallon. Accordingly, selection of information which allows for a comparison of the hypothesised categories on this features is very likely to discriminate between them. Although such a selection is likely to involve more cognitive effort, it is also likely to result in more cognitive effects.
The experiments to be presented in this chapter will examine whether such a trade-off between cognitive effect and cognitive effort is at work on the pseudodiagnosticity task. The first experiment to be reported should be thought of as an exploratory study, whilst the second experiment will follow up on the results of the first.

2.2 EXPERIMENT 1

2.2.1 Introduction

2.2.1.1 Manipulating the probability of the features

The rationale for this, the first experiment of the thesis, is simple. If subjects are sensitive to the probabilities with which events occur in their environment (as both the previous analysis, and work using the selection task suggest), then manipulating the probability of the evidential features with which subjects are presented in the pseudodiagnosticity task should affect the rates at which they select the remaining pieces of evidence. This may be illustrated with reference to Mynatt et al's car problem. The features which they presented to their subjects were doing over 25 mpg and no major mechanical problems in the previous two years. Both of these features are quite common amongst the general population of cars. An alternative analysis is necessary, however, in the case where subjects are told that their sister's car possesses a rare attribute and a common attribute (e.g. it has leather upholstery and does over 25 mpg) and are then told the percentage of one type of car which possesses the rare feature. This alternative analysis is necessary because extrapolating from the work on subjective probability, described in Chapter 1, such a manipulation would be predicted to have profound effects on the additional piece of information which subjects perceive to be relevant. Subjects receiving an initial piece of information about a rare feature should tend to focus on information concerning \( p(D_1/H_y) \) whilst subjects who receive initial information about a common feature should focus on information concerning \( p(D_2/H_x) \). This is a direct consequence of subjects' sensitivity to the base rates of features of objects in their environment. The lower
the probability of a feature, the less likely it is that any object or category of objects will possess it. Accordingly, if one knows that members of a hypothesised category possess a rare feature in common with an object to be categorised, information about the rate at which members of the alternative hypothesised category posses that feature is likely to be informative.

Unfortunately, the position is not quite as simple as the foregoing would suggest. Mynatt, Doherty and Dragan (1993) found that the value of \( p(D_1/H_x) \) was important in determining evidence selections amongst their subjects. In Experiment 1 of their study all subjects were told that \( p(D_1/H_x) \) was 0.65 whereas in their second experiment all subjects were told that \( p(D_1/H_x) \) was 0.35. On their inference problems (where subjects were told that the car belonged to their sister) the majority of subjects selected a further piece of information concerning \( H_x \) when \( p(D_1/H_x) \) was set at 0.65. However, on these same problems when \( p(D_1/H_x) \) was set at 0.35, the majority of subjects selected information about \( p(D_1/H_y) \). Mynatt et al explain their results by claiming that when \( p(D_1/H_x) \) fell below 0.5 subjects reasoned pseudodiagnostically that \( H_x \) was probably false and shifted the focus of working memory to \( H_y \). Importantly, Mynatt et al fail to take into consideration the base rate of the initial evidential feature. In contrast to Mynatt and his colleagues it is predicted here that an interaction between \( p(D_1/H_x) \) and the perceived base rate of \( D_1 \) should be observed. The specific details of this prediction are presented below. Firstly, it is predicted that where the subjective probability of \( D_1 \) is high the rate at which information about \( p(D_1/H_y) \) is selected will increase as \( p(D_1/H_x) \) decreases. Secondly, it is predicted that where the subjective probability of \( D_1 \) is low, the rate at which information about \( p(D_1/H_y) \) is selected will increase as \( p(D_1/H_x) \) increases. Experiment 1, which attempts to test these predictions, will shortly be described. The second aim of this experiment must firstly be discussed.

2.2.1.2 The use of diagnostic and pseudodiagnostic information

As mentioned in Chapter 1, where the pseudodiagnosticity task was first introduced, there has been interest in the use to which subjects put diagnostic information.
There have been three studies (described on pgs. 35 - 36) which have investigated the use to which subjects put diagnostic information (Doherty et al 1979; Beyth-Marom and Fischhoff, 1983; Ofir, 1988). Unfortunately, there are two problems with all three of these studies. Firstly, in all cases subjects were given base rate evidence in addition to information about the likelihood ratios. Accordingly, there is no direct evidence about how subjects' use diagnostic information. It is entirely possible that in the absence of base rate information subjects will use the likelihood evidence in a more appropriate manner. Secondly, and much more importantly, the emphasis in all of these studies has been on what subjects do with normative evidence. This is despite the fact that the majority of subjects, when given the opportunity, do not choose that normative evidence. This problem obviously relates to the problem of the status of the normative theory in research on decision making and reasoning (see Baron, 1988 for a discussion of this issue). Evans (1991) has characterised standard logic as a yardstick against which subjects' performance on a variety of reasoning tasks may be evaluated.

This view would also seem to be prevalent in much of the decision making literature and leaves unaddressed the question of what subjects actually do with the "incorrect" information which they tend to select on evidence choice tasks. As one of the major goals of this thesis is to investigate precisely how subjects revise their beliefs in a set of hypotheses in the light of evidence relevant to these hypotheses, the second aim of the first study described here will be an investigation of exactly that question. Subjects will be given an initial piece of evidence and then asked to choose one further piece of information to help them decide between two hypotheses and, upon receipt of that evidence, will be asked to use it to judge the relative likelihood of the hypotheses. It should be pointed out at this stage that such a procedure is not standard on the pseudodiagnosticity task. Although Tweney, Doherty, Mynatt and Schiavo (1979) did ask subjects to use the information which they received, Tweney et al's task was more complicated than the one which will be used here. Accordingly, it is very difficult to draw conclusions from their results. However, even though the task used by Mynatt, Doherty and Dragan (1993) was much simpler than that used by Tweney et al, the former experimenters did not require their subjects to put the information which they received to any use.
The main focus of this chapter remains the factors affecting subjects' selection of information for the testing of hypotheses. Accordingly subjects' use of the information which they receive in the course of the first experiment will not be discussed at great length. It will be returned to in Chapters 3 and 4 where it will be used as the basis for a more thorough investigation of the phenomena of belief revision. However, even at this stage it is worth pointing out there exists much experimental and theoretical work (which will be reviewed in Chapter 3) which suggests that subjects will revise their beliefs about a set of hypotheses in the light of non-diagnostic information. For example, the literature on impression formation is based on a non-normative paradigm where the issue of diagnosticity is rarely discussed (e.g. Asch, 1946; Jacobson and Anderson, 1965). There also exist several experimental examples of subjects revising their beliefs in the light of non-diagnostic information. Troutman and Shanteau (1977) provided subjects with samples from one of two boxes of beads. One box had 70 red, 30 white and 50 blue beads. The other box had 30 red, 70 white and 50 blue beads. Subjects' task was to state the subjective probability that the samples which they were shown came from the first or the second box. Because both boxes contained equal numbers of blue beads, a sample consisting of blue beads was uninformative. Yet subjects significantly revised their subjective probability estimates when they were given blue beads, in the direction of less extreme judgements.

This finding leads to the prediction that even those subjects who select pseudodiagnostic information will use the information they receive to form beliefs which favour one or other of the hypotheses. Other than this prediction no other predictions were made about subjects' overall levels of belief in this experiment.

2.2.2. Method

Subjects: the subjects were 144 students of the Business School at the University of Plymouth.
Materials: each subject was presented with a booklet containing a set of instructions and one problem. The instructions were as follows:

Accompanying these instructions is a decision problem. On the first page there is a brief description of a situation and you are asked to select a single piece of information in order to make a decision between two alternatives. You may feel that you would like to have more than one piece of information, but please pick only one. We are interested in which piece you think would be most useful in helping you make a decision, even though ideally more information might be helpful.

Once you have marked the piece of information you think would be most useful in helping you make a decision, please remove the sticker which corresponds to that piece of information. It is important that you remove only the sticker corresponding to the piece of information which you have chosen.

On the next page is a scale designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that one of the decision alternatives is true and the other to complete certainty that the second alternative is true. Please mark the point on the line that corresponds to your beliefs about the alternatives. Remember that the greater your confidence in an alternative, the closer your mark should be to that alternative.

It is very important to read the problem carefully and to think about it before picking the piece of information you think would be most useful in helping you make a decision. Take your time and consider your choice before you respond. Likewise, it is very important that you think carefully before deciding which point on the line best describes your beliefs about the likelihood of each
decision alternative. If you have any questions at any time, please raise your hand and the experimenter will help you.

The structure of the problems used was similar to that of Mynatt et al (1993). Each subject was given a problem such as the following:

Your friend is an engineer and works for a large construction company. It's either company X or company Y, but you can't remember which. You do remember that he earns over £14,000 a year and that he drives a company car.

You have the following pieces of information:

(a) 65% of engineers working for company X earn over £14,000 a year.

Three additional pieces of information are also available:

(b) The percentage of engineers working for company Y that earn over £14,000 a year.

(c) The percentage of engineers working for company X that drive a company car.

(d) The percentage of engineers working for company Y that drive a company car.

Assuming that you could find out only ONE of these pieces of information (b, c, or d), which would you want in order to help you decide what company your friend works for?

There were three other such problems concerning a friend who has just moved house, a friend who is staying in a hotel, and a friend who has just started university. Examples of each of these problems are presented in Appendix 1.

Underneath each problem was the following statement:
Once you have marked your answer please peel back the sticker which corresponds to that answer. Underneath is the piece of information which you have selected. Now turn to the next page where you will find a scale designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend works for company X and the other to complete certainty that he works for company Y. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to that alternative.

This statement was, of course, changed slightly for each scenario. On the next page was a line 100 mm. long with one end marked "certain.......X" and the other marked "certain.......Y".

**Design:** as there are two parts to this experiment, the design of each part will be outlined separately. The first part of the study involved subjects' selection of evidence for the evaluation of the hypotheses present in the scenarios and had a 2x3x4 between subjects design. The first factor was Rarity and involved manipulating the feature about which the initial piece of evidence was given. Subjects received an initial piece of evidence about either a rare or a common feature of the object which they had been asked to categorise. The features used are given in Table 2.1.

<table>
<thead>
<tr>
<th>Problem Content</th>
<th>Rare Feature</th>
<th>Common Feature</th>
<th>2nd Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineer</td>
<td>earns £50,000 p.a. +</td>
<td>earns £14,000 p.a. +</td>
<td>company car</td>
</tr>
<tr>
<td>House</td>
<td>swimming pool</td>
<td>garden</td>
<td>garage</td>
</tr>
<tr>
<td>Hotel Room</td>
<td>costs £165/night</td>
<td>costs £35/night</td>
<td>en-suite bathroom</td>
</tr>
<tr>
<td>Friend starting</td>
<td>2As &amp;1B at A-level</td>
<td>1B &amp;2Cs at A-level</td>
<td>must take up a foreign</td>
</tr>
<tr>
<td>University</td>
<td></td>
<td></td>
<td>language</td>
</tr>
</tbody>
</table>

*Table 2.1: Features used in Experiment 1. The 2nd feature column refers to the feature kept constant across all conditions.*
The second factor was Likelihood and had three levels. The Likelihood factor involved manipulating the level at which the first piece of evidence said the feature was present. Subjects were told that 65% or 95% of instances of category X possessed the feature or were simply told that instances of category X possessed the feature. In this latter case subjects were told that the available evidence was whether instances of X or Y possessed the features. It was assumed that informing subjects that instances of category X possessed the initial feature corresponded to telling them that 100% of instances of that category possessed the feature. Accordingly, this will be referred to as the 100% level of the Likelihood variable. The third factor was problem content. As has already been stated, there were four different contents used.

The second part to this experiment concerned the use to which subjects put the evidence which they received. This part of the study had a 2x3x2x2x4 design. Alongside the factors involved in the first part of the study, there were a further two factors involved in the second part of the study. The first of these was Feedback. This involve manipulating the actual value of the evidence which subjects had selected. In all conditions this was printed beside the appropriate piece of information and was covered by an opaque sticker. Subjects were asked to mark the piece of evidence they had chosen and then to remove the corresponding sticker. In the 65% and 95% conditions subjects received evidence that was either positive or negative relative to category X. Positive evidence was supplied by having 25%, 75%, and 25% under the stickers corresponding to b, c and d respectively (a brief re-examination of the example given above will make this clear). Negative evidence was supplied by giving the likelihoods in the order 75%, 25%, 75%. Those subjects who were told that instances of category X possessed the initial feature received feedback in a categorical yes/no form. Once again the sign of the evidence was determined by giving either yes or no as feedback to each of the possible evidential items.

The second, additional, factor involved in the second part of the study was Choice. This refers to whether subjects had selected a diagnostic or a pseudodiagnostic piece of evidence. The design of both parts of the experiment is summarised in Table 2.2.
Task Independent Variables Levels

Choice of Information
1. RARITY 2 - RARE, COMMON
2. LIKELIHOOD 3 - 65%, 95%, 100%
3. PROBLEM CONTENT 4 - see Table 2.1

Use of Information
1, 2 and 3
4. FEEDBACK 2 - POSITIVE, NEGATIVE
5. CHOICE 2 - DIAGNOSTIC, PSEUDODIAGNOSTIC

Table 2.2: Independent and Dependent variables involved in Experiment 1

Procedure: subjects were run in two groups, one of 97 and the other of 47. The experiment was run in class with subjects being assigned a problem randomly.

2.2.3. Results and Interpretation

Due to a typing error, the results of subjects on the house problem were invalidated. Of the remaining 108 subjects, one failed to provide a confidence rating, and only eight chose evidence about \( P(D_2/H_Y) \). Accordingly, the following results are those of the 99 subjects who completed the engineer, hotel, and university problems and who chose to receive evidence about \( P(D_1/H_Y) \) or \( P(D_2/H_X) \).

2.2.3.1 Information Choice

There was no effect of problem content on subjects choice of information \( (X^2 = 1.09, \text{df} = 2, p > .25) \). Thirty three percent of subjects who received the engineer problem chose \( P(D_1/H_Y) \) versus 61% who chose \( P(D_2/H_X) \). The equivalent figures for the university and hotel problems were forty two percent versus fifty percent, and forty three percent versus forty nine percent respectively.

The effect of Rarity and Likelihood on subjects' selection of evidence was analysed using log linear models. This is a technique designed to overcome the problems inherent in
multi dimensional contingency tables (see Howell, 1992 or Kennedy, 1983 for comprehensive introductions to the subject). The simplest explanation of the approach is that it involves comparing ever simpler models of the data until the most parsimonious model is found. In this case the most parsimonious model was that of a main effect for Likelihood (likelihood ratio chi square = 2.63, df = 3, p > .45). Although there were other models which did not predict frequencies which were significantly different to the observed frequencies, this model is appropriate because all other potential models were more complex, and the model chosen did not differ significantly from these alternative models. Subjects' response frequencies, broken down by Rarity and Likelihood are given in Table 2.3.

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Common</th>
<th></th>
<th>Rare</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>65%</td>
<td>11</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>95%</td>
<td>9</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>100%</td>
<td>4</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>26</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2.3: Response frequencies, broken down by Rarity and Likelihood, for Experiment 1.

A chi-square was used to test for significant differences amongst the three levels of the Likelihood variable. The results are represented in Figure 2.1 whilst selection frequencies for the entire experiment are presented in Appendix 2. Overall, this was significant ($X^2 = 14.67, df = 2, p < .001$). Follow up chi squares revealed significant differences between the 95% level and the 100% level ($X^2 = 8.77, df = 1, p < .005$), and between the 65% level and the 100% level ($X^2 = 13.89, df = 1, p < .001$). There was a significant difference in frequency of diagnostic evidence selections (i.e. selection of $P(D_1/H_y)$) due to the Likelihood variable ($X^2 = 9, df = 2, p < .025$). This difference was significant between the 95% level and the 100% level ($X^2 = 6.5, df = 1, p < .025$), and between the 65% level and the 100% level ($X^2 = 6.5, df = 1, p < .025$). There was a corresponding significant difference in pseudodiagnostic selections (i.e. selection of $P(D_2/H_x)$) due to feature level ($X^2 = 8.19, df = 2, p < .025$). The only significant difference was that between the 65% level and the 100% level ($X^2 = 4.9, df = 1, p < .05$).
The only level of the Likelihood variable at which there was a significant difference between frequency of diagnostic and pseudodiagnostic selections was the 100% level ($X^2 = 15.13$, df = 1, $p < .001$).

The only significant difference in subjects' choice of information was caused by the Likelihood manipulation. Simply telling subjects that instances of category X possessed the first feature greatly increased the frequency with which subjects chose the pseudodiagnostic information. There are two possible explanations for this. The first of these is that giving subjects this information caused them to focus on the category to which that information related to the exclusion of the other category. This fits with the Mynatt et al (1993) account of subjects' pattern of evidence selection.

![Figure 2.1: The percentages of subjects choosing diagnostic and pseudodiagnostic information in Experiment 1 broken down by the level at which the first feature was said to be present amongst instances of X.](image)

However, there is an alternative view of this result. Telling subjects that instances of category X possessed the first feature, and giving them further information in terms of whether instances of categories X and Y possessed the remaining features, changed the structure of the task. This may be seen in Table 2.4. This change in the structure of the task is due to the fact that telling subjects whether instances of a category possess a feature...
amounts to giving them propositional information. This propositional information allows subjects, in certain cases, to be certain in their decision between the alternatives. As may be seen from Table 2.4, a NO outcome from any of the possible pieces of information allows subjects to be certain about which of the categories the object belongs to. This is because a NO outcome means that instances of the category to which that piece of evidence relates do not possess at least one of the features possessed by the target object. For two out of the three pieces of information (whether instances of Y possess either of the features) a YES outcome leads to the situation where each of the categories possesses at least one of the features. Thus, in these two cases, subjects may be said to possess non-diagnostic information.

<table>
<thead>
<tr>
<th>Initial Evidence</th>
<th>Remaining Evidence</th>
<th>Possible Outcomes</th>
<th>Results if Outcome</th>
<th>Results if Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Instances of category X possess initial feature.</td>
<td>B. Whether instances of category Y possess the initial feature.</td>
<td>YES or NO</td>
<td>Uncertainty. Both categories possess the initial feature.</td>
<td>Certainty. The object must be an instance of X.</td>
</tr>
<tr>
<td>C. Whether instances of category X possess the second feature.</td>
<td>YES or NO</td>
<td>Dependent on inference about the diagnosticity of the evidence.</td>
<td>Certainty. The object must be an instance of Y.</td>
<td></td>
</tr>
<tr>
<td>D. Whether instances of category Y possess the second feature.</td>
<td>YES or NO</td>
<td>Uncertainty. Each of the categories possesses at least one feature.</td>
<td>Certainty. The object must be an instance of X.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4: Initial evidence, remaining possible evidence, possible outcomes, and the results of those outcomes when subjects are given propositional information on the pseudodiagnosticity task.

However, information about whether instances of X possess the second feature (the choice overwhelmingly made by subjects in this experiment) is the most interesting case. As stated above, a NO outcome allows subjects to infer that the target object is an instance of Y. A YES outcome, on the other hand, means that instances of category X possess both of the features possessed by the target object. Logically this does not allow subjects to infer
that the target object is an instance of X. It may still be the case that all of the possible evidence is non-diagnostic and both of the features are possessed by instances of both categories. However, if subjects expect the information which they are given to be useful then, having been told that instances of category X possess the first feature, the optimal strategy is to select evidence about whether instances of category X possess the second feature. If instances of category X do possess the second feature then the target instance must belong to category X (assuming that the experimenter is not playing a trick and the evidence is diagnostic). If instances of category X do not possess the second feature then the target object must belong to the category Y. It may be the case then, that subjects who received evidence in a propositional format made the pragmatic inference that the evidence which they would receive was diagnostic and, based on this inference and a consideration of all possible selections and their contingent outcomes, decided to select information about whether instances of X possessed the second feature.

That this second explanation is the case is unlikely for at least one a priori reason. For each of the possible evidence selections there are two alternative outcomes. This means that subjects must reason through six possible outcomes (see Table 2.4) in order to arrive at the correct selection. Even if it is the case that most subjects will not consider whether members of category Y possess D₂, there are still four possible outcomes for subjects to reason through. Much of the work earlier mentioned has provided evidence that subjects, when reasoning, often fail to consider alternatives. This view is especially prevalent in the work of Johnson-Laird and colleagues (e.g. Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991). It seems much more likely therefore that subjects reason through the alternatives involved one by one, starting at the point to which the information in the problem directs them, and settle on a choice once they have an alternative which they feel will give them some information.

2.2.3.2 Confidence Ratings

It will be remembered from the Introduction to this experiment that it differs from much previous work using the pseudodiagnosticity paradigm. This difference is due to the
subjects in this experiment being required to make use of the evidence which they received in the course of the task. They were asked to make use of that evidence by providing a rating of their confidence in the hypotheses in the light of all the evidence which they had available to them. In this section of the chapter subjects' confidence ratings will be analysed.

As was pointed out in the previous section, telling subjects that instances of a category possess a feature (as was the case in the 100% level of the Likelihood manipulation) changed the structure of the task. Accordingly, the confidence ratings for the 65% and 95% levels of the Likelihood variable will be analysed separately to those for the 100% level (i.e. where subjects were given evidence in propositional form).

2.2.3.2.a Confidence ratings with probabilistic evidence

As was the case with the previous analysis of subjects' information selection, the confidence ratings of those subjects who attempted the house problem were also invalidated by a typing error. As before, this leaves 108 subjects who produced valid results. However, as only those subjects in the 65% and 95% levels of the Likelihood condition will be included in this first analysis, there will be a maximum of 72 subjects whose results will be analysed. Of these 72 subjects who attempted the engineer, university or hotel problems, five selected information about $P(D_2/H_y)$. Accordingly these subjects will not be included.

It will be remembered that subjects were asked to express their confidence on a 100mm line. At one end of the line was labelled "certain...X" and at the other end "certain...Y". For the purposes of the analysis which follows subjects' markings on the line were converted to scores on a 100 point scale ranging from 1 to 100. Each point on this scale corresponds to one millimetre on the line and the higher is a subject's score on this scale the more confident she is that the X hypothesis is the case.
Due to the reduced number of subjects, two separate analyses had to be carried out. The first of these was a one way Anova on the effect of problem content. This was not significant (F(2, 64) < 1). The mean confidence ratings for the engineer, university, and hotel problems were 58 (S.D. = 26), 66 (S.D. = 26), and 56 (S.D. = 24) respectively. The summary table from this analysis is presented in Appendix 2.

The main analysis performed was a 2x2x2x2 between subjects Anova to examine the effect of Rarity, Likelihood, Feedback and Choice (i.e. whether subjects chose diagnostic or pseudodiagnostic evidence) on confidence ratings. Choice is not a true independent variable as the level to which subjects were assigned (diagnostic or pseudodiagnostic choice) was not determined in advance by the experimenter but was instead determined by subjects' information selection. Accordingly, the effects of the Choice variable may only be interpreted in a correlational, rather than a causal, sense. The summary table from this analysis is presented in Appendix 2 as is a full table of means and standard deviations.

The analysis produced four effects of interest, three of which were significant. The main effect of Likelihood was highly significant (F(1, 51) = 10.29, p < .003). The mean confidence rating for subjects in the 65% condition was 49 (S.D. = 24), and for subjects in the 95% condition mean confidence was 70 (S.D. =23). This result is interesting because

Figure 2.2: The significant two way interaction between Likelihood and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1.
there was no equivalent significant main effect for Feedback. Subjects seemed to be sensitive to the level at which the feature is said to be present in the first piece of evidence but not in the second. This is despite the fact that differences between levels was more pronounced in the second piece of evidence (i.e. 75% vs. 25%) than in the first (65% vs. 95%). There was also a significant interaction between Likelihood and Choice (F(1, 51) = 7.09, p < .02). This interaction is represented in Figure 2.2 and the means and standard deviations involved are given in Table 2.5.

Tests for simple main effects showed that the effect of Likelihood was highly significant when subjects chose diagnostic information (F(1, 36) = 21.23, p < .01), but not when subjects chose pseudodiagnostic information (F(1, 29) < 1).

<table>
<thead>
<tr>
<th>Choice</th>
<th>Likelihood 65%</th>
<th>Likelihood 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic</td>
<td>43 (22) N = 20</td>
<td>76 (11) N = 17</td>
</tr>
<tr>
<td>Pseudodiagnostic</td>
<td>59 (24) N = 13</td>
<td>65 (30) N = 17</td>
</tr>
</tbody>
</table>

**Table 2.5:** Means, standard deviations (in parentheses) and number of subjects involved in the significant two-way interaction between Likelihood and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1.

This result may be explained in terms of the relationship between the confidence ratings produced by those subjects who chose diagnostic evidence and the correct Bayesian probabilities given that evidence. Although it is not claimed that subjects should be expected to have produced confidence ratings which closely approximated Bayesian probabilities, it should be the case that changes in subjects' confidence due to differences in both Likelihood and Feedback mirror those prescribed by Bayes theorem. As may be seen from Table 2.6, this was not the case.

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Feedback</th>
<th>Confidence</th>
<th>Bayesian Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>65%</td>
<td>Positive (25%)</td>
<td>51</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>Negative (75%)</td>
<td>34</td>
<td>.46</td>
</tr>
<tr>
<td>95%</td>
<td>Positive (25%)</td>
<td>72</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>Negative (75%)</td>
<td>78</td>
<td>.56</td>
</tr>
</tbody>
</table>

**Table 2.6:** The confidence ratings of subjects who chose diagnostic evidence in Experiment 1, broken down by Likelihood and Feedback, and the corresponding Bayesian probabilities given the evidence.
The first thing which an inspection of Table 2.6 reveals is that the significant two-way interaction between Likelihood and Choice may be attributed to the relative under-confidence of subjects in the 65% conditions with both types of feedback (with positive feedback, actual rating of 51 vs. Bayesian probability of .72; with negative feedback, actual rating of 34 vs. Bayesian probability of .46), and to the overconfidence of subjects in the 95% conditions who received negative feedback (mean rating of 78 vs. Bayesian probability of .56). The second thing which is apparent is that mean ratings are not even in the normatively correct order of magnitude. It would seem that even with the minimum amount of information required to make a judgement (i.e. the likelihood ratio), subjects do not follow Bayes theorem in their confidence ratings.

The third effect of interest was the marginally non-significant interaction between Rarity and Choice (F(1, 51) = 3.79, p < .06). The means involved in this interaction are given in Table 2.7. Once again tests for simple main effects were carried out which revealed that the effect of Rarity was significant when subjects chose the pseudodiagnostic evidence (F(1, 29) = 5.37, p < .05), but not when they made the diagnostic choice (F(1, 36) < 1).

<table>
<thead>
<tr>
<th>Choice</th>
<th>Common</th>
<th>Rare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic</td>
<td>55 (25) N = 20</td>
<td>61 (23) N = 17</td>
</tr>
<tr>
<td>Pseudodiagnostic</td>
<td>72 (27) N = 14</td>
<td>54 (25) N = 16</td>
</tr>
</tbody>
</table>

Table 2.7: The means, standard deviations (in parentheses) and subject numbers involved in the interaction between Rarity and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment

This interaction may be explained in terms of the significant three way interaction between Rarity, Feedback, and Choice (F(1, 51) = 4.59, p < .04). The means and standard deviations involved in this interaction are given in Table 2.8, whilst the interaction is presented in Figure 2.4. Tests for simple interaction effects revealed a significant interaction between Rarity and Feedback when subjects chose pseudodiagnostic information (F(1, 26) = 6.03, p < .025) but not when they chose diagnostic information.
The significant half of the three way interaction was tested for simple main effects. Rarity had a significant effect when subjects received positive feedback ($F(1, 14) = 10.99, p < .01$) but not when they received negative feedback ($F(1, 14) < 1$).

Although unexpected, this finding, on the surface seems quite clear. The Rarity manipulation had a significant effect only when subjects chose pseudodiagnostic evidence for which they received positive feedback. The mean confidence rating of subjects who, having already received information about a common feature, chose pseudodiagnostic

<table>
<thead>
<tr>
<th></th>
<th>Diagnostic Selection</th>
<th>Pseudodiagnostic Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive Fback</td>
<td>Negative Fback</td>
</tr>
<tr>
<td>Common</td>
<td>54</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>25.73</td>
<td>25.98</td>
</tr>
<tr>
<td></td>
<td>(N = 8)</td>
<td>(N = 12)</td>
</tr>
<tr>
<td>Rare</td>
<td>65</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>23.05</td>
<td>23.22</td>
</tr>
<tr>
<td></td>
<td>(N = 9)</td>
<td>(N = 8)</td>
</tr>
</tbody>
</table>

Table 2.8: Means (in bold), standard deviations and subject numbers involved in the significant three way interaction between Rarity, Feedback and Choice from the analysis of subjects' confidence ratings in the 65% and 95% conditions of Experiment 1.
evidence, and received positive feedback, was 83. Subjects who chose exactly the same evidence, and received the same feedback, but who had first received information about a rare feature produced a mean confidence rating of 45.

One possible explanation for this is that the effect of being told that either 65% or 95% of instances of category X possessed a rare feature was to cause subjects to expect a very high level of members of that same category to possess the common feature. When that expectation was not met, subjects' confidence in category X was diminished. Those subjects who had initially received information about a common feature did not have those expectations and so the effect of the second piece of evidence was not to confound expectations. The problem with this explanation is that it does not explain the pattern of results for those subjects who had chosen pseudodiagnostic evidence and received negative feedback. Here there was no effect of the Rarity manipulation. It may be the case, however, that these subjects, faced with two overtly contradictory pieces of evidence, resolved the contradiction by discounting one of those pieces of evidence. If this is the case then the mean confidence of those subjects in the 95% conditions should be higher than for those subjects in the 65% conditions. This is the observed pattern for those subjects who received initial information about a rare feature (65 vs. 54) but not for those subjects who received information about a common feature (56 vs. 60). However, cell sizes at this level of analysis are too small for these means to be considered reliable. It should be remembered, however, that one of Ofir's (1988) findings was that subjects will use likelihood information as a means of resolving inconsistencies between the base rates. With only two pieces of inconsistent information available it is likely that subjects resorted to some characteristic of one of the evidential items to resolve that inconsistency. As for the expectation account, it will be developed and tested in Chapters 3 and 4. Therefore, the above results will be discussed again at the start of Chapter 4.

2.2.3.2. b Confidence ratings with propositional evidence

As was stated in the previous section, it was decided to analyse the confidence ratings of those subjects in the 65% and 95% Likelihood conditions separately from the
ratings of those subjects in the 100% Likelihood condition. This is because subjects in the 100% condition were given evidence in a propositional format. This change in evidential format, from probabilistic to propositional, also changed the logical structure of the task. Accordingly, the analysis of the confidence ratings of those subjects in the 100% Likelihood condition will be presented in this section of the chapter.

One third of all of the subjects who participated in this experiment received evidence in propositional form. This means that there was a maximum of 36 subjects whose confidence ratings could have been included in this analysis. Of these 36 subjects, four failed to provide a confidence rating, and only five selected information concerning $p(D_2/H_y)$. It was decided to exclude these subjects from the analysis and instead, to analyse the results of the 27 subjects who, having been told that instances of category X possessed the first feature, selected the pseudodiagnostic evidence.

Once again, two analyses had to be carried out (again, summary tables and tables of means are to be found in Appendix 2). The first of these was a one way Anova on the effect of problem content. This was not significant. The mean confidence rating for the engineer problem was 57 (S.D. = 30.42, N = 11), for the university problem 57 (S.D. = 28.84, N = 9), and for the hotel problem 52 (S.D. = 39.23, N = 7).

The second analysis was a 2x2 between subjects Anova to examine the effect of Rarity and Feedback on confidence ratings. For subjects given information in propositional format who chose the pseudodiagnostic evidence, positive feedback was achieved by telling subjects that instances of category X did possess the second evidential feature. Negative feedback consisted of informing subjects that instances of category X did not possess the second feature.

The effect of Feedback was found to be significant ($F(1, 23) = 5.73, p < .03$). The mean confidence rating of those subjects who received positive feedback was 68 (S.D. = 16.50, N = 14) whilst the mean for those who received negative feedback was 41 (S.D. = 37.95, N = 13). Neither the main effect of Rarity ($F(1, 23) = 0.36, p > .5$) nor the
interaction between Rarity and Feedback ($F(1, 23) = 1.38, p > .25$) were found to be significant.

Of most interest here is the significant main effect of Feedback on subjects' mean confidence ratings. Although Feedback had a significant effect on subjects' relative confidence in the hypotheses, the means involved in the interaction do not suggest that subjects understood the task. Remember that changing the format in which subjects received the information from probabilistic to propositional also changed the logical structure of the task. Accordingly, those subjects who, having chosen a further piece of information relevant to the hypothesis about which they already possessed some evidence, received negative feedback, could be confident that the target object was not an instance of category X. The mean confidence rating for these subjects does not suggest that subjects realised this. Those subjects who received negative feedback had a mean rating of 41, far short of complete confidence that Y was the case.

Likewise, those subjects who received positive feedback had a mean confidence of 68, once again far short of complete confidence that the target object was an instance of category X. Of course, as mentioned in a previous section, the structure of the task justified subjects' lack of confidence in category X. Although subjects who received positive feedback subsequent to a pseudodiagnostic selection knew that instances of category X possessed both of the evidential features, they did not know anything about instances of category Y. Thus subjects may have felt unsure that even if they had all of the information available to them they would have been able to decide between the alternatives. For example, instances of both X and Y categories might have turned out to possess both evidential features. It was earlier suggested that a partial cause of subjects' overwhelming tendency to select the pseudodiagnostic evidence when told that instances of category X possess the first evidential feature may have been the making of the pragmatic inference that the evidence which was available on the task was diagnostic. In other words, subjects may have inferred that instances of both X and Y categories did not possess both of the evidential features. However, the confidence ratings of subjects who received positive feedback does not suggest that this is the case. If subjects had inferred
that only one of the categories possessed both of the features, then information that instances of category X possessed both of the features should have resulted in near certainty that the target object was an instance of category X. This was not the case.

2.2.4 Discussion

The results of Experiment 1 are both surprising and intriguing. Although no evidence was found to suggest that subjects are sensitive to the rarity of evidential features when they select further evidence to test hypotheses, changing the format (from probabilistic to propositional - as was achieved with the Likelihood manipulation) of the information which they received had a significant effect on their patterns of information selection. It seems likely that the significant effect of changing the format in which subjects receive information is due either to the change in the logical structure of the task or to the manner in which evidence in a propositional format directs subjects' attention.

However, the failure to find a significant result due to the probabilities of the evidential features about which subjects received information need not necessarily be taken as evidence against the view that subjective probability plays a role in determining the information which subjects choose to test hypotheses. It should be remembered that Experiment 1 was an exploratory study and, as such, it contains several weaknesses. The first of these is that no pre-test was carried out to establish whether subjective probabilities of the evidential features used were as expected. Accordingly, such a pre-test was carried out before Experiment 2 was run. This pre-test will be described in the next section. The second weakness of this experiment is that it lacks statistical power. It may therefore be argued that the failure to find a significant result due to the Rarity manipulation may have been the result of too few subjects rather than to the non existence of such an effect. For this reason, Experiment 2 will employ a simpler design with greater numbers of subjects in each cell of the design.

As stated previously, the interesting result from the information selection data collected in Experiment 1 was the significant effect which the Likelihood manipulation had
on subjects' pattern of evidence choice. There are two explanations for the observed patterns of results. Subjects initially told that instances of category X possessed the first evidential feature (i.e. given evidence in propositional form) may have selected further information about category X in overwhelming numbers firstly, because they appreciated the logical structure of the task, and secondly, because they assumed that the information available was useful. Having reasoned through all of the possible choices and outcomes they then selected the pseudodiagnostic piece of evidence - the single piece of evidence which they felt would allow them to decide between the hypotheses. A negative outcome from this selection (i.e. being told that instances of category X did not possess the second evidential feature) would enable subjects to conclude that the target object was an instance of category Y. A positive outcome (i.e. instances of category X possess the second, as well as the first, evidential feature), on the other hand, would enable subjects to conclude that the target object was an instance of category X, but only if the additional assumption that instances of both categories did not possess both features, had been made.

As pointed out previously, such an explanation is unlikely to account for the observed pattern of results. This is because it is unlikely that most subjects would, or indeed could, engage in the kind of hypothetical thinking necessary to arrive at the correct selection in such a way. There is also the fact that subjects' ratings of their confidence in the hypotheses after receipt of both pieces of evidence does not suggest that they understood the logical structure of the task. Although a significant main effect of Feedback was found, the means involved suggest that subjects did not comprehend the task. For example, those subjects who received negative feedback were entitled to be completely certain that the target instance belonged to category Y. In fact, the mean confidence rating for subjects who received negative feedback was 41. On this basis alone it seems unlikely that the latter explanation for subjects' behaviour is correct.

Whether subjects made the pragmatic assumption that the possible evidence was useful is a moot point. It may be argued that subjects could not be sure that any of the information which they might receive would enable them to discriminate between the hypotheses. Such a line of argument would claim that subjects only had access to the
information they were given and may not have assumed that this information was
diagnostic. Accordingly, subjects who chose a further piece of information about category
X and received positive feedback may have considered the possibility that information
about category Y would make membership of category Y as probable as membership of
category X. This counter-argument is weakened, however, by the fact that it does not seem
to apply to the results of those subjects who received information in a probabilistic format.
Although those subjects were equally entitled to doubt the diagnosticity of the evidence
which they received, their confidence scores do not seem to reflect such a doubt.
Nevertheless one of the goals of Experiment 2 will be to rule out the above counter-
argument.

As stated in the introduction to Experiment 1, the main focus of the experiment was
subjects' evidence selections. Therefore the results on the confidence rating measure will
not be discussed in great detail here. The main interest of the confidence ratings from those
subjects given information in a propositional format is in relation to possible
interpretations of subjects' evidence selections on those problems. The confidence ratings
produced by subjects who received probabilistic information are of far greater interest.
Firstly, these results suggest that subjects are not good Bayesians. The significant two way
interaction between Likelihood and Choice was caused by subjects who chose diagnostic
information being unable to produce confidence judgements which mirrored the
prescriptions of Bayes theorem.

It would seem that expectations created by the initial piece of information are
important in determining subjects' level of confidence in the hypotheses. If this is the case
it suggests that subjects are sensitive not only to the probability with which objects or
events occur in their environment but also to the extent to which the occurrence of one
event may be predicted from the occurrence of another. Once again, however, caution must
be exercised in interpreting these results. As mentioned above, there is no independent
measure of the subjective probabilities of the evidential features used in this experiment
and the experiment lacks power. Neither was any direct measure of subjects' expectations
nor of their confidence in the hypotheses taken after receipt of the first piece of evidence.
Experiment 3, which will be described in Chapter 4, will constitute an attempt to further investigate the issues raised here. Before attempting to clarify the issues raised by subjects’ confidence ratings on Experiment 1 it was decided to further investigate some of the questions concerning information selection which were also raised by this experiment.

2.3. EXPERIMENT 2

2.3.1 Introduction

As pointed out in the discussion of Experiment 1 there were several problems with both the design of the study and the interpretation of subjects' information selections. Firstly, because there were not enough subjects in the first experiment it is impossible to draw the conclusion that subjects were insensitive to the probability of occurrence of the features about which they received evidence even though no significant main effect was found for the Probability variable. Secondly, because the finding of no significant main effect for the Rarity manipulation is questionable, no conclusions can be drawn about the strategies subjects used in deciding which piece of evidence to select. As pointed out in the discussion of Experiment 1, subjects may have reasoned through all of the possibilities until they found one which would allow them to be certain in their decision between the alternatives or they may have considered only that piece of information to which their attention was directed by the problem content. The experiment to be described here will attempt to provide clarification of both these issues.

It was also pointed out in the discussion of Experiment 1 that the evidential features used in that experiment had not been pre tested for their perceived probability of occurrence. In this second experiment it was decided to use problems containing both rare and common features which had been pre-tested. The pre-test will be described in the next section.
Essentially this experiment was a large scale replication of one part of Experiment 1. All subjects were given information about the hypotheses in propositional form. That is, all subjects were told initially that instances of category X possessed one of the evidential features and were presented with further possible information concerning whether instances of the categories possessed the evidential features. As before, subjects were given an initial piece of information about either a common or a rare feature of an object to be categorised and asked to select a further piece of information in order to help them decide between hypothesised categories. Also as before, the target object was said to possess two features. This meant that there were three remaining pieces of information from which subjects had to choose. It was pointed out in the discussion of the previous experiment that a possible confound existed in the structure of the propositional form of the task. This confound is that although the selection of an initial piece of information about the hypothesis concerning which subjects are given an initial piece of information may allow them to discriminate between the hypotheses, they have no guarantee that this will be case. In other words, both of the hypothesised categories may possess both of the features, or both may possess just one of the features. Although the suggestion that this confound might have affected the results of Experiment 1 was deemed unlikely, it was controlled for in this experiment by use of a Hint manipulation. Accordingly, half of the subjects received an additional piece of information in the instructions stating that there was at least one piece of information available which would enable them to decide between the alternatives. This Hint manipulation controlled for the effects of subjects' assumptions about the diagnosticity of the evidence available to them. It was pointed out in the discussion of Experiment 1 that subjects' selection of further evidence may be dependent on the pragmatic assumption that the evidence available is diagnostic. Subjects who selected a further piece of evidence concerning the hypothesis about which they already possessed some information may have done so because they assumed that this piece of information was the only one which would allow them to decide between the alternatives with certainty, regardless of outcome. On the other hand, some subjects may not have made this assumption and so selected information about instances of category Y. In the Hint condition of this experiment subjects' selections were not dependent on any such pragmatic assumption. This is because telling subjects that there was at least one piece of information
available to them which would allow them to decide between the alternatives with certainty was in effect, telling them that the evidence was diagnostic.

2.3.2 Pre-test

The aim of this pre test was to select the features for four problems to be used in Experiment 2. Two of the problem contexts used in Experiment 1 were retained (the house and engineer problems) and a further two were added. The first of these new contexts involved a car which the subject's sister was said to have bought whilst the second concerned a villa where the subject's parents were said to have stayed.

Subjects: the pre test was run in two stages. In the first stage, materials for the house, engineer and car problems were tested. 30 subjects took part in this stage of the pre test, 6 of whom were male and 24 female. Their mean age was 27 with the youngest subject being 19 and the oldest 43. In the second stage of the pre test 24 subjects rated materials for the Spanish villa problem. 18 of these subjects were male whilst 6 were female. Their mean age was 24 with the oldest subject being 37 and the youngest 19. All subjects were undergraduate students at the University of Plymouth and were paid for their participation.

Materials: all subjects received a handout which comprised an instruction sheet and four sheets each of which contained a context and a list of features to be rated. Subjects in the first stage of the pre test rated features for the house, engineer and car problems and an additional problem whilst subjects in the second stage of the pre test rated materials for the Spanish villa problem and an additional three problems (none of these additional problems were used in the experiments described in this thesis).

For all subjects the instructions were as follows:

*On the following pages you will find a series of three questions, each of which has eight subsections. Each question concerns one particular class of things.*
The subsections ask you to estimate how often, in every ten thousand occurrences, an item will have a particular feature.

So, for example, you may be asked:

Out of every 10,000 students how many would you expect to

(a) own a car

(b) have a part-time job

If you feel that the answer to the first part of the question is 1,550, then you should write this in the space provided.

Please try to answer all of the questions. While it is unlikely that you will know the exact answer to any of the questions, we are interested in your best guess. This does not mean that we want you to put down the first number that pops into your head. Your knowledge of each of the things in the questions should enable you to give sensible answers in most cases!

If you have any questions, please raise your hand and the experimenter will help you.

At the top of each rating sheet was a question in the following form:

Out of every 10,000 holiday villas in Spain, how many would you expect to

This was followed by a list of features for which subjects had to give frequency estimates. Each of these lists contained between 8 and 10 features. A copy of these materials is given in Appendix 4.

Procedure: subjects were run in groups of between 5 and 10 in experimental sessions which had no time limit.
Results: mean estimations and standard deviations for all of the features rated are given in Appendix 4. Mean estimations (in the form of percentages) for the features used in Experiment 2 are given in Table 2.9.

<table>
<thead>
<tr>
<th>PROBLEM CONTENT</th>
<th>Feature Rated</th>
<th>Mean Estimations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineer</td>
<td>earns over £60,000 p.a. earns</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>£25,000 per annum</td>
<td>49%</td>
</tr>
<tr>
<td></td>
<td>drives a company car</td>
<td>62%</td>
</tr>
<tr>
<td>House</td>
<td>has a swimming pool</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>has a garden</td>
<td>81%</td>
</tr>
<tr>
<td></td>
<td>has a garage</td>
<td>56%</td>
</tr>
<tr>
<td>Car</td>
<td>top speed of 165 m.p.h. + has a</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>radio</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>has four doors</td>
<td>58%</td>
</tr>
<tr>
<td>Spanish Villa</td>
<td>costs £1000 per week</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>costs £150 per week</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td>built in last twenty years</td>
<td>84%</td>
</tr>
</tbody>
</table>

Table 2.9: Problem contents and mean frequency estimations for the features used in Experiment 2.

These features were chosen because they fell at either end of the distribution of subjects' estimations. However, it should be borne in mind that it was impossible to control completely for the perceived frequency of the features used. For example, although subjects expected only 12% of civil engineers to earn over £60,000 per annum, this was over three times greater than the number of houses they expected to have a swimming pool. Likewise, some features were regarded as being more common than others.

2.3.3 Method

Subjects: 96 subjects participated in this experiment as part of their course requirements. Subjects' mean age was 21. The youngest subject was 18 whilst the oldest was 36. 28 of the subjects were male and 68 were female. In addition to receiving course credit subjects were also paid for participation.
Materials: each subject received a handout which comprised an instruction sheet and four problems. The instructions given to half of the subjects were as follows:

Accompanying these instructions is a series of four decision problems. Each consists of a description of a situation. The following is an example of the type of situation we have used:

Your friend has just bought a new television. You can't remember whether it's a model X or a model Y but you do remember that it has teletext and a remote control.

Next you will be given a piece of information about the situation. These pieces of information are given in terms of the question you would ask to receive the information and answers which actually tell you what you want to know. For the example situation above you might be told

**Question A**

whether model X televisions have remote controls

Answer: Yes

Now you know that model X televisions do have remote controls.

Following each piece of information you will be given a list of the three further questions you could ask with their possible answers. You will be asked to rate the potential usefulness of each question in helping you to decide between the X and Y alternatives present in the description of the situation. If you feel that knowing the answer to a particular question would be extremely helpful in deciding between the two alternatives, you should place a mark at the "OF GREAT USE" end of the rating scale. If you feel that knowing the answer to a question would not be helpful in deciding between the two alternatives you should place your mark at the "OF LITTLE USE" end of the scale. Remember! The more or less useful that you think a question might be, the closer to the appropriate end of the scale you should place your mark.
After you have filled in the rating scales you will be asked to select the question which you think would be most helpful in deciding between the X and Y alternatives. You may feel that you would like to ask more than one question, but please pick only one. We are interested in which question you think would be most useful in helping you make a decision, even though ideally receiving the answer to more than one question might be useful.

It is very important that you read the problems carefully and think about them before filling in the rating scales or picking the question you think would be most useful in deciding between the two alternatives. Take your time and consider your choice before you respond.

If you have any questions at any time, please raise your hand and the experimenter will help you.

As mentioned in the Introduction to this experiment, it was decided to include a Hint manipulation in the study. This manipulation was intended to control for the possible effects of subjects' pragmatic assumptions about the overall informativeness of the evidence on their pattern of information selections. Accordingly, the remaining subjects received the same instructions as above but were also told:

N.B. For each of the four problems there is at least one question you can ask which, regardless of the answer to that question, will allow you to be certain in your decision between the alternatives.

The problems used were similar to those used in Experiment 1. In this experiment however, subjects were asked to rate the usefulness of each piece of information before making a choice. Subjects received no feedback once they had chosen the piece of information which they considered to be most useful. However, as in Experiment 1, a Rarity manipulation was employed in this experiment. Below is an example of one of the problems used in the Common condition.

Your sister bought a new car in 1988. You can't remember whether it's a model X or a model Y but you do remember that it has four doors and a radio.
We have already asked the following question for you and have given you the answer:

**Question A**

whether model X cars bought in 1988 have a radio

*Answer: YES*

Three additional questions are possible which we have listed below along with their possible answers:

**Question B**

whether model Y cars bought in 1988 have a radio

*Possible Answers: YES or NO*

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

**Question C**

whether model X cars bought in 1988 have four doors

*Possible Answers: YES or NO*

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

**Question D**

whether model Y cars bought in 1988 have four doors

*Possible Answers: YES or NO*

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

Assuming that you could discover the answer to only one of these questions (B, C, or D), which would you ask in order to help you decide which model car your sister drives?
Subjects at the Common level of the Rarity manipulation received problems such as the one above. Subjects at the Rare level received problems which were exactly the same except that the initial piece of information which they received concerned a rare feature (in this case a top speed of over 165 mph). In summary, as with Experiment 1, there were two levels of the Rarity variable in this experiment. The first of these is the Rare level, where subjects were given an initial piece of information concerning a rare feature of the class of object to be categorised. The second level of the Rarity variable was the Common level where subjects were given an initial piece of information concerning a common feature of the target object. The four problem contents, with the features which were used, are shown in Table 2.10. Examples of each of the problems are given in Appendix 1. As may be seen, there were two versions of each problem. When crossed with the instructional manipulation this gives four separate conditions: Common with and without a hint, and Rare with and without a hint.

The initial piece of information which subjects received always concerned the X alternative. In problems with two common features the initial information always concerned the same common feature, whilst in problems with one rare feature the initial piece of information always concerned that rare feature. The other possible pieces of information - whether instances of Y possessed the first feature, whether instances of X possessed the second common feature, whether instances of Y possessed the second common feature - will be referred to as B, C, and D respectively. To make this change in notation easier to follow it may be useful to remember that B corresponds to \( p(D_1/H_Y) \) on the standard version of the task, whilst C corresponds to \( p(D_2/H_X) \) and D corresponds to \( p(D_2/H_Y) \). Unfortunately, this change in notation is made necessary by the fact subjects were given information in propositional rather than probabilistic form in this experiment. This change in information format also renders the terms diagnostic and pseudodiagnostic unsuitable as it changes the structure of the task.
A Latin square design was used to control for the order in which B, C, and D appeared (and hence, the order in which their potential usefulness was rated). Four subjects in each condition were presented with one of the six possible orderings of B, C, and D for each problem. The order in which subjects received the problems was randomised.

<table>
<thead>
<tr>
<th>FEATURE PAIRS</th>
<th>Common</th>
<th>Rare</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROBLEM CONTENT</td>
<td>Engine</td>
<td>House</td>
</tr>
<tr>
<td></td>
<td>25,000 per annum</td>
<td>earns over £60,000 p.a.</td>
</tr>
<tr>
<td></td>
<td>drives a company car</td>
<td>drives a company car</td>
</tr>
<tr>
<td></td>
<td>has a garden</td>
<td>has a swimming pool</td>
</tr>
<tr>
<td></td>
<td>has a garage</td>
<td>has a garage</td>
</tr>
<tr>
<td></td>
<td>has a radio</td>
<td>top speed of 165 m.p.h. +</td>
</tr>
<tr>
<td></td>
<td>has four doors</td>
<td>has four doors</td>
</tr>
<tr>
<td></td>
<td>costs £150 per week</td>
<td>costs £1000 per week</td>
</tr>
<tr>
<td></td>
<td>built in the last twenty years</td>
<td>built in the last twenty years</td>
</tr>
</tbody>
</table>

Table 2.10: Problem contents and features used in Experiment 2

Procedure: subjects were run in groups of between 4 and 12 in experimental sessions which had no time limit.

Design: there were two dependent variables in this study. The first of these was the overall total of C choices (i.e. a further piece of information relevant to the hypothesis about which subjects already possessed some information) which subjects made throughout the experiment. This part of the study had a 2x2 between subjects design. The factors were Hint (the presence or absence of a hint), and Rarity (two common features or one common and one rare).

The second dependent variable was subjects' ratings of the usefulness of each possible piece of information (Usefulness Rating). This had a 2 x 2 x 3 x 4 mixed design. Once again, the between subjects factors were Hint (the presence or absence of a hint) and Rarity (common or rare initial feature). The within subjects factors were Item Rated (B, C or D) and problem content. The design of the entire experiment is summarised in Table 2.11.
Two subjects in the no hint/rare and common features condition failed to complete at least one of their four problems. Accordingly, they will not be included in the following analyses.

**2.3.4. Evidence selections**

The first dependent variable in this experiment was the total number of C (i.e., a further piece of information about instances of category X) choices made by subjects across the four problem contents. Before presenting an analysis of the results on this measure, however, overall patterns of information selection for the experiment will be discussed. The percentage of subjects selecting each of the three possible pieces of information, broken down by the Rarity and Hint manipulations, are shown in Table 2.12, whilst a frequency table for the entire experiment is given in Appendix 2.

Across all conditions, the choice frequencies for problem contents were very similar. On the engineer problem, 31% of selections were of B, 53% of C, and 11% of D. On the house problem the equivalent percentages were 34%, 62%, and 4%. On the car problem they were 33%, 61%, and 6%, and on the villa problem they were 39%, 52%, and...
As expected, item C was the most highly selected piece of information, followed by item B and then item C. The numbers of item B, item C, and item D choices were tabulated for each subject across the four problem contents. The mean numbers of item choices, broken down by the Rarity and Hint manipulations, are given in Table 2.13. A 2x2 between subjects Anova was carried out on the mean number of item C choices in all four between subjects conditions (a summary table for the Anova is given in Appendix 2). A significant main effect was found for Rarity (F(1, 90) = 4.31, p < .05). Neither the main effect of the Hint manipulation (F(1, 90) = 1.00, p > .3), nor the interaction between Hint and Rarity (F(1, 90) = 2.01, p > .15) were found to be significant.

<table>
<thead>
<tr>
<th></th>
<th>COMMON</th>
<th>RARE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>HINT</td>
<td>25%</td>
<td>75%</td>
</tr>
<tr>
<td>NO HINT</td>
<td>35%</td>
<td>56%</td>
</tr>
</tbody>
</table>

Table 2.12: Percentages of subjects choosing each item broken down by instructional manipulation and feature type for Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>COMMON</th>
<th>RARE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>HINT</td>
<td>1.00</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>1.44</td>
<td>1.44</td>
</tr>
<tr>
<td>NO HINT</td>
<td>1.42</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>1.42</td>
</tr>
</tbody>
</table>

Table 2.13: Mean item selections (with standard deviations in bold) broken down by features and instructional manipulation from Experiment 2

2.3.4.2. Usefulness Ratings

The second dependent measure involved in this experiment was Usefulness Ratings. Subjects rated the usefulness of each piece of information by placing a mark on a line 100 millimetres long one end of which was labelled "Of little use" whilst the other end
was labelled "Of great use". Subjects' confidence ratings were converted to scores on a 100 unit scale, running from 1 to 100, where each millimetre on the line corresponded to one unit on the scale. The higher a subject's score for a particular piece of information the more useful they rated that piece of information to be.

A 2x2x3x4 mixed design Anova was used to analyse the results on this measure. The between subjects factors were Rarity and Hint, whilst the within subjects factors were Item Rated and problem content. A full Anova table for this analysis is given in Appendix 2 as is a table of all of the means and standard deviations. The significant main effect for Item Rated (F(2, 182) = 18.25, p < .001) was as expected and follow up Tukey HSDs revealed significant differences between all three means (p < .005 in all cases with the exception of the difference between item B and item D where p < .02). The mean usefulness rating for item B was 64 (S. D. = 22). For item C the mean rating was 71 (S. D. = 18), and for item D the mean rating was 57 (S. D. = 18).

The significant main effect of problem content (F(3, 273) = 2.70, p < .05) was not expected however. The mean rating of usefulness for items on the engineer, house, car and villa problems were 63 (S. D. = 17), 67 (S. D. = 17), 63 (S. D. = 16), and 63 (S. D. = 18) respectively. Tukey HSDs revealed no significant differences between any of these means. The only other significant result revealed by this analysis was the interaction between Item Rated and problem content (F(6, 546) = 2.48, p < .025). Tukey HSDs revealed no significant differences between the mean rating for each item across problem contents. The significance of this interaction, as Figure 2.4 suggests, was due to the fact that, for the car problem, the mean usefulness rating for B was significantly lower than for C but not significantly lower than for D. This is in contrast to the pattern present amongst the other contents where the difference between B and C ratings was not significant but the differences between both B and D and C and D were significant.

Of the remaining higher order interactions, three approached significance. The least interesting of these is the four way interaction between Rarity, Hint, Item Rated and problem content (F(6, 540) = 1.96, p < .07). Much more interesting are the two way
interaction between Hint and Item Rated (F(2, 180) = 2.70, p = .07), and the three way interaction between Rarity, Hint and Item Rated (F(2, 180) = 2.48, p < .09). The two

Figure 2A: The interaction between the Item Rated and problem content from the analysis of subjects’ usefulness ratings in Experiment 2.

Table 2.14: Means (in bold), standard deviations (in parentheses) and significant differences between means for the interaction between Hint and Item Rated from the analysis of subjects' usefulness ratings in Experiment 2.

<table>
<thead>
<tr>
<th>Hint Manipulation</th>
<th>Item Rated</th>
<th>1 68</th>
<th>2 70</th>
<th>3 59</th>
<th>4 60</th>
<th>5 73</th>
<th>6 55</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Hint</td>
<td>B</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hint</td>
<td>B</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tukey HSDs revealed that there were no significant differences between the item ratings across conditions. In other words, there were no significant differences between any of the four mean ratings for item B, no significant differences between any of the four mean ratings for item C and no significant differences between any of the ratings for item D.
Figure 2.5: The interaction between Hint and Item Rated from the analysis of subjects' usefulness ratings in Experiment 2.

Figure 2.6: The interaction between Rarity, Hint and Item Rated from the analysis of subjects' usefulness ratings in Experiment 2.

<table>
<thead>
<tr>
<th>ITEM:</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO HINT</td>
<td>68 (20)</td>
<td>70 (17)</td>
<td>63 (19)*</td>
</tr>
<tr>
<td>HINT</td>
<td>69 (23)</td>
<td>66 (18)</td>
<td>70 (15)*</td>
</tr>
</tbody>
</table>

Table 2.15: The means and standard deviations (in parentheses) involved in the interaction between Rarity, Hint, and the Item Rated from the analysis of subjects' usefulness ratings in Experiment 2. (* denotes a mean significantly lower than the rating for common/hint item C.)
2.3.5 Discussion

The results of this experiment are, on the whole, as expected. Firstly, the pattern of evidence selection was as predicted. Overall, item C was the most frequently selected, followed by item B, with item D a poor third. Also as predicted, the Rarity manipulation had a significant effect on the mean number of item C choices which subjects made. Subjects who received an initial piece of information concerning a rare feature were significantly less likely to select a further piece of information relevant to category X than were subjects who received an initial piece of information concerning a common feature. This is interpreted as support for the claim that subjects' background knowledge about subjective probability, in some cases, informs their selection of information for the evaluation of hypotheses. The fact that item C was chosen significantly less often when subjects were given an initial piece of information about a rare feature also suggests that subjects will attend to, and therefore select, the information which is most likely to have the greatest effect on their beliefs (see the next section for a demonstration of this).

Interestingly, this conclusion is in contrast to the recent claim by Evans and Over (1996) that subjects will select information which is relevant to the hypothesis which they currently consider most likely. This does not seem to be the case here, as subjects given information concerning a rare feature (presumably offering stronger support for the hypothesis to which it related than did receipt of information concerning a common feature) were significantly less likely to select further information concerning that hypothesis than were subjects given an initial piece of information about a common feature.

Subjects given an initial piece of information about a rare feature seem to have inferred that knowing that instances of the category Y do not possess that rare feature will tell them that the target instance is a member of the focal category. It is uncertain as to whether they also inferred that knowing that instances of category Y possess the rare feature would not allow them to make a decision. However, given what is known about subjects' general failure to consider alternatives in many tasks, it is a reasonably safe assumption that subjects are satisfied with choosing B (p(D1/Hy) on the standard version
of the pseudodiagnosticity task) because firstly, they consider it likely to provide them with discriminatory information and secondly, they have failed to consider alternative possible pieces of information.

The Hint manipulation had no effect on subjects' pattern of information selection. However, the results of the second dependent measure - Usefulness Ratings - are not as clear-cut in this regard as are the results on the Choice measure. Overall, the mean usefulness ratings for each of the evidential items correspond to the frequency with which subjects chose those items. However, the Hint manipulation seems to have had a greater influence here than it did on the choice data. This is evidenced by both the interaction between Hint and Item Rated and the interaction between Probability, Hint and Item Rated. Nevertheless, there were no significant differences found between the individual means for any of the items rated in either of the interactions i.e. there were no significant differences between ratings for B, no significant differences between ratings for C etc. Thus, it may be inferred that although the Hint manipulation had an effect on the differences between ratings of the items within each condition, it did not have a significant effect across conditions. The manipulation merely seems to have affected how subjects used the scale to discriminate between the usefulness of the alternative pieces of information.

In a sense this experiment asks more questions than it answers, and certainly there is much more work to be done before we understand exactly what kinds of reasoning subjects are engaging in on the task. However, the experiment was not designed to examine the minutiae of the task. Rather, it was run as a demonstration of how subjects' knowledge about the likelihood that features will occur in their environment affects the information they select to evaluate hypotheses. In demonstrating this, the experiment has also shown that subjects in Experiment 1 were not reasoning through all of the consequences of selecting each piece of available information. If they had done so on this experiment, then the Rarity manipulation would not have had a significant effect on subjects' evidence selections. As Sperber, Cara and Girotto (1995) suggest, subjects seem to have inferred directly testable consequences from the initial information they were given. The order in which they made those inferences seems to have depended on what
was being focused upon at the time. Thus, subjects who received an initial piece of information concerning a common feature predominantly chose a further piece of information about category X, not because it would enable them to discriminate between the hypotheses, but because of their initial representation of the problem. In other words, their initial representation was anchored (see Evans and Over, 1996b for a discussion of anchoring and representation) around the X hypothesis thus making further information about the X hypothesis seem relevant. The relevance of these results to the discussion of the literature in Chapter I will now be discussed.

2.4. GENERAL DISCUSSION

The results of these experiments provide support for the model of information selection discussed in the conclusion of Chapter 1. It was argued that a model which captures both general relevance theoretic principles and specific findings concerning sources of information available to the subject (such as knowledge of subjective probabilities) is required. The first claim which may be made about these experiments is that they demonstrate the importance of subjects' background knowledge concerning the probability of events in their environment to their selection of information for the evaluation of hypotheses.

It is not clear to what extent the increase in selection of information about category Y amongst subjects who received an initial piece of information about a rare feature is due to their initial representation of the task (to talk about subjects' initial representation of the task is equivalent to discuss what parts of the scenario subjects are initially focusing on). Evans and Over (1996b) suggest that a similar tendency amongst the subjects in Mynatt, Doherty and Dragan's (1993) study was due to the manner in which subjects represented the task. They argue that by foregrounding the features, Mynatt et al caused subjects to construct an initial representation based upon the feature contained in the initial piece of information which subjects received. Although this may be the case here, subjects were not told that the decision between the alternatives had any utility for them. Neither were the
features foregrounded. Exactly the same problem structure was used with both common and rare features. It may be the case that the rare feature was fore-grounded because of its rarity. However, it is also plausible that subjects who received an initial piece of information about a rare feature chose further information about that feature even if it was not initially fore-grounded. The information which subjects expected to gain from the selection of evidence concerning category Y may have warranted the extra processing required to select that evidence. Although the design of Experiment 2 does not allow a decision as to which of these readings of the results is the correct one, both readings sit with a model which balances general principles of cognitive economy against specific sources of knowledge available to subjects in background knowledge.

It is possible to formalise the information which subjects might have expected to gain from each of the pieces of evidence available to them on the version of the pseudodiagnosticity task used in Experiment 2. This may be done by conceptualising the task as one of making a decision about which piece of evidence to select. Next, the formula used in normative decision theory to calculate the subjective expected utility of any choice may be adapted for an analysis of the task. This formula is typically expressed as follows:

\[ SEU = \sum_i s_i U_i \]  

Eq. 2.1

where \( s_i \) refers to the probability of the \( i \)th possible outcome of the choice and \( U_i \) represents the utility of that outcome, and where \( i \) ranges over a finite set of mutually exclusive and exhaustive outcomes. As applied to the version of the pseudodiagnosticity task used in Experiment 2, the formula may be used to calculate the information which subjects expect the selection of a particular piece of evidence to yield.

Applying the formula to the task used in Experiment 2 is a relatively easy affair. For ease of exposition the structure of the task is represented in Figure 2.7. For every possible piece of evidence there are only two possible mutually exclusive and exhaustive outcomes: the category to which the piece of evidence
relates either possesses the particular feature or it doesn't. From the pre-test for Experiment 2, the frequency with which subjects expect the various features to be possessed by objects in the world is also known. Although there is some debate in the literature about the psychological equivalence of frequency and probability information (see Gigerenzer and Hoffrage, 1995; Cosmides and Tooby, 1996; Kahneman and Tversky, 1996), in this case it is likely that subjects' frequency judgements provide a reasonably good estimation of the perceived probability of the features used. The approximate probabilities from the pre-test are given in Table 2.16. All that remains is a calculation of the utilities associated with each particular outcome. Although this is difficult to do in practice,

![Diagram](image)

**Figure 2.7:** The structure of the variant of the pseudodiagnosticity paradigm used in Experiment 2.

<table>
<thead>
<tr>
<th>Initial Evidence</th>
<th>Remaining Possible Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instances of X possess D1</td>
<td>B - whether instances of Y possess D1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.16:** The probabilities and utilities associated with each possible outcome from the task used in Experiment 2.
an examination of Table 2.16 reveals that whilst it is not possible to assign exact utility values in every case, it is possible to say something about the likely relationships which must exist between the utilities of some of the possible outcomes on the task. It may also be seen that some outcomes have been assigned a utility of 0. This is true of YES outcomes for both of the pieces of information which concern category Y. YES outcomes in these instances simply tell the subject that instances of both categories possess at least one of the features possessed by the target object, and so do nothing to reduce her uncertainty.

It will be seen from Table 2.16 that a utility of m has been assigned to a NO outcome for each of the possible pieces of information. In terms of the logical structure of the task, a NO outcome allows subjects to be completely certain which of the categories the target object belongs to. It could be argued therefore, that all of the NO outcomes should be assigned an infinite utility. There are two problems with this. Firstly, as Evans and Over (1996b) point out, it is unlikely to be the case that no uncertainty remains concerning a set of hypotheses after an observation or set of observations has been carried out. For this reason alone it is more satisfactory to think of hypotheses which have been verified or falsified as having probabilities close to 1 or 0. Secondly, if the measure of the utility of an outcome which is used happens to be the absolute log likelihood ratio (see Eq. 2.2), as is advised by both Evans and Over (1996a&b) and Laming (1996), problems arise when one starts to talk of verification or falsification in absolute terms.

\[
\text{Epistemic Utility of Outcome} = \text{ABS} \left[ \log \frac{\text{Prob}(E/H)}{\text{Prob}(E/not-H)} \right]
\]

Eq. 2.2

As may be seen from Eq. 2.2, the utility of any outcome is determined by the posterior probability of the hypotheses given the evidence. The absolute value of the log likelihood ratio is taken to ensure that the informativeness of any piece of information will be symmetrical i.e. epistemic utility will increase as the probabilities move away from 0.5 regardless of whether the outcome favours H or not-H. The problems referred to above arise if it is claimed that the evidence discriminates absolutely between H and not-H i.e. if it is claimed that H has been absolutely verified by E. For this claim to be made,
Prob(E/H) must equal 1, Prob(E/not-H) must equal 0, and the likelihood ratio becomes, very inconveniently, infinity.

Such a problem does not arise if a value of m is assigned to the NO outcome from each of the possible pieces of evidence. This value, m, should be understood as a number close to infinity. Likewise, assigning a value of 0 to two of the outcomes in Table 2.16 is, strictly speaking, implausible. These outcomes should have utilities close to, but not equal to, 0. For ease of exposition, however, they will be assigned utilities of 0.

The only utility value remaining to be discussed is the utility of n assigned to a YES outcome arising from a request for further information concerning whether instances of category X possess the second evidential feature. A YES outcome means that instances of category X possess both of the features possessed by the target object but does not tell us anything about instances of category Y. The precise value of n will depend on several things. Firstly, it will depend on whether subjects even consider instances of category Y. It is likely that the less attention that is paid to category Y, the greater the value of n becomes. In the cases where subjects do attend to instances of category Y, the value of n may sometimes depend on whether the pragmatic assumption that the evidence available is diagnostic has been made by the subject. Imagine the subject considering whether she should choose to find out whether instances of category X possess the second evidential feature. If instances of X do turn out to possess that feature, what would she be able to infer? If she thinks that it is likely that instances of both categories will turn out to possess both of the features, then her present uncertainty will be reduced very little by a YES outcome. On the other hand, if she thinks it unlikely that the experimenter would play such a trick on his subjects (in other words, if she assumes that the evidence available is diagnostic) then her uncertainty will be greatly reduced by a YES outcome. If the experimenter is not playing a trick on her, then it is extremely unlikely that instances of both categories will possess both of the features, so the target object is very likely to be an instance of category X.
The non-success of the Hint manipulation in Experiment 2 seems to suggest that the value assigned to n does not depend on subjects' pragmatic assumptions about the diagnosticity of the evidence. If subjects' selection of information was dependent on such assumptions then the effect of the Rarity manipulation should have been less marked in the condition where subjects were told that at least one piece of evidence, regardless of its outcome would lead to complete certainty in deciding between the alternatives. The only piece of information which fits this bill is further information about whether instances of category X possess the second feature. However, when given an initial piece of information concerning a rare feature, subjects who received a hint were just as likely to select information concerning whether instances of category Y possessed the rare feature as were subjects who did not receive the hint. If subjects were considering the problem at a level where assumptions about the diagnosticity of the evidence was important then an increase in selection of further information about category X would have been expected. However, it is also possible that the failure of the Hint manipulation may have been due to the wording used. Subjects were told that regardless of outcome there was, at least, one piece of information which would allow them to decide between the alternatives. Subjects may have understood this to refer to the logic of the problems rather than the actual answers to their questions. This understanding of the hint may have contradicted subjects' understanding of the logical structure of the task, causing them to ignore the hint.

When the expected information yield for each possible piece of evidence is calculated (see Table 2.17), the importance of the value of n may be seen.

<table>
<thead>
<tr>
<th>Choice</th>
<th>Initial Feature</th>
<th>Expected Information Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>whether instances of Y possess the initial feature</td>
<td>Common</td>
<td>.3m</td>
</tr>
<tr>
<td></td>
<td>Rare</td>
<td>.9m</td>
</tr>
<tr>
<td>whether instances of X possess the second feature</td>
<td>Common</td>
<td>.7n + .3m</td>
</tr>
<tr>
<td></td>
<td>Rare</td>
<td>.7n + .3m</td>
</tr>
<tr>
<td>whether instances of Y possess the second feature</td>
<td>Common</td>
<td>.3m</td>
</tr>
<tr>
<td></td>
<td>Rare</td>
<td>.3m</td>
</tr>
</tbody>
</table>

*Table 2.17: Expected information yield for each possible piece of information, broken down by Probability, from Experiment 2.*
From Table 2.17 it may be seen that when subjects are given an initial piece of evidence concerning a common feature, asking whether instances of X possess the second evidential feature is the best option. This, of course, depends on the utility assigned to \( n \) being greater than 0. However, as was argued earlier, the value of \( n \) will depend on the extent to which subjects attend to category Y. Previous research (Doherty, Mynatt, Tweney and Schiavo, 1979; Mynatt, Dragan and Doherty, 1993), and the theorising of Evans and Over (1996b), suggests that on the standard version of the pseudodiagnosticity paradigm, subjects fail to consider the alternative hypothesis present on the task. It is likely, therefore, that \( n \) will be assigned a relatively high value, thus leading to a predominance of subjects asking for further information concerning category X.

In the case where subjects receive an initial piece of information concerning a rare feature, however, the expected information yield associated with asking whether instances of category Y also possess that initial feature is three times as great as when the initial piece of evidence concerns a common feature. The expected information yield from asking whether instances of category X also possess the second feature is, normatively, exactly the same with an initial piece of information about a rare or a common feature. However, in the case where the initial piece of evidence concerns a rare feature, the utility assigned to a YES outcome in response to a request for further information concerning category X is likely to be lower than when the initial piece of evidence concerns a common feature. This is because the value assigned to \( n \) depends on the extent to which category Y is attended to. Because the presence of a rare feature will cause subjects to attend to category Y, the value assigned to \( n \) will be less than in the case where subjects receive an initial piece of information about a common feature. This explains why there were significantly fewer selections of further information about category X amongst subjects at the rare level of the Rarity manipulation than there were amongst subjects at the common level.

What is most interesting about the above model of the task used in Experiment 2 is the premium which it places on psychological factors in the assignment of specific utilities to particular outcomes. This interdependence between probabilistic and psychological factors captures exactly the point made in Chapter 1 about the varied possible sources of
relevance. The value which subjects assign to n (the utility of a YES outcome in response to a request for information about whether instances of X possess the second feature) is dependent on the extent to which subjects represent, and consider, category Y. As Evans and Over (1996b) argue, it is impossible to give a purely probabilistic account of either the information from background knowledge, or the information contained in the problem, which subjects will attend to. Instead, a recognition of the complex interplay between notions of cognitive effect (i.e. expected information yield) and cognitive effort (i.e. a limited representational system) is required.

The experiments contained in this chapter contain an additional point of interest. Although Experiment 1 did not provide evidence concerning the importance of subjective probability for the selection of information to evaluate hypotheses, it did demonstrate the importance of the format in which subjects receive that information. In so doing Experiment 1 has produced a new task for the study of hypothesis testing. As was stated in Chapter 1, the experimental study of hypothesis testing has centred around two tasks: the Wason selection task and the pseudodiagnosticity task. These tasks differ in the normative model which may applied to them. The Wason selection task may be characterised as a conditional reasoning problem (e.g. Wason, 1966) or as a probabilistic decision making task (Evans, Over and Manktelow, 1993; Oaksford and Chater, 1994; Evans and Over, 1996), whilst the pseudodiagnosticity task has always been characterised as a test of subjects' understanding of Bayes theorem. However, by changing the format in which subjects received information on the pseudodiagnosticity problem, the entire structure of the task was changed. This new version of the task cannot, in any sense, be thought of as requiring subjects to select the piece of information which would enable them to apply Bayes theorem. Instead, it demands that subjects infer the consequences of selecting each possible piece of evidence. The analysis of the new task, presented here, suggests that a similar analysis of the standard version of the pseudodiagnosticity task might profitably be undertaken. It is also very similar to Sperber, Cara and Girotto's (1995) recent characterisation of the selection task in that it claims that subjects will attempt to infer the consequences of information selections until they have reached a conclusion which they regard as relevant (or satisfactory). This suggests that fruitful use of this new task may be
made in investigating the relative importance of cognitive effect and cognitive effort to information selection for hypothesis evaluation.

Experiment 2 constituted an attempt to clarify the issues of characterisation raised by Experiment 1. In the earlier experiment, subjects overwhelmingly selected further information about the hypothesis for which they already possessed some evidence. However, their confidence ratings upon receipt of that second piece of information suggested that they were not reasoning through all of the consequences involved in selecting each piece of evidence. Likewise, the significant effect of the Rarity manipulation in Experiment 2 also suggests that subjects were not reasoning through all of possible outcomes for each piece of evidence. This, in itself, is support for the notion that cognitive economy is vital in information selection. Human beings are cognitive misers. This fact must be used to temper any account given of the manner in which subjects use a specific source of information such as background knowledge about subjective probability.

2.5 SUMMARY

The results of two experiments using the pseudodiagnosticity paradigm, and a variant of that paradigm, were described in this chapter. The main aim of Experiment 1 was a demonstration of the effect of subjective probability on subjects' information selection for hypothesis evaluation. Although the experiment failed to provide such a demonstration, it did provide evidence that changing the format in which subjects receive information on the task, from probabilistic to propositional, dramatically alters both subjects' behaviour on the task, and the structure of the task itself. Subjects were also asked to use the information which they were given on the task. The results were unforeseen and suggested that subjects' expectations about unseen evidence are crucial in determining the effect of that evidence on their confidence in the hypotheses to which it relates.

The second experiment, in demonstrating an effect of subjective probability on subjects' evidence selection, also suggested that the failure of Experiment 1 to do so was
due to its lack of statistical power. All of the problems used in Experiment 2 gave subjects information in a propositional format. The significant effect of subjective probability in this experiment suggested that the effect of the Likelihood manipulation in Experiment 1 was not due to the change in the logical structure of the task which propositional information entailed. Instead it was argued that subjects' pattern of information selection on the task may be explained in terms of their initial representation of the problems and a trade off between effort and effect.

The results of these studies were interpreted as support for a model of information selection incorporating general relevance theoretic principles and a specific account of the sources of information which subjects have available in background memory. In support of this argument a model of the expected information yield for each possible selection of further evidence was proposed. According to this model, subjects' selection of information in Experiment 2 may be thought of as being dependent not just on the probabilistic structure of the task, but also on the constrained representational resources of the human information processing system. Finally, it was pointed out that the propositional version of the pseudodiagnosticity task constitutes a new instrument for the study of the selection of information for hypothesis evaluation.
CHAPTER 3 - THE ORDER EFFECTS PARADIGM AND MODELS OF BELIEF REVISION

3.1. INTRODUCTION

It is now time to turn to the second question asked in the Introduction to this thesis. How is evidence used to evaluate hypotheses? This question has rarely been addressed in the reasoning literature. However it is addressed in both the decision making and information integration literatures. This chapter will review and critically discuss evidence from both of these literatures. In the final chapter of this thesis it will be argued that many of the basic mechanisms which are held to underlie information selection also underlie how information is used to decide between alternative hypotheses. However, in the information integration and belief revision literatures very different accounts of the process have been proposed. These accounts are based, to a large extent, on the order effects paradigm. As the account given in Chapter 2, of the way in which subjects used the information they received, will be tested in the next chapter using an order effects paradigm, that paradigm will be reviewed here. Both recency and primacy effects will be discussed, as well as the many attempts in the literature to account for their existence. This will be followed by a discussion of the models which have been put forward in the information integration literature to account for belief revision generally, and order effects specifically. It will be argued that there are several a priori reasons for doubting the adequacy of all of these models.

Chapters 4 and 5 will contain the details of six experiments using the order effects paradigm. As mentioned above, the first of these experiments will use an order effects paradigm to test the account given in Chapter 2 of the results from Experiment 1. Subsequent experiments will use the same paradigm to demonstrate the importance of pragmatics to any account of belief revision. Accordingly, a review such as the one which will be attempted in this chapter is necessary before presenting the details of those experiments.
One very important point concerning the status of Bayesian approaches to belief revision must be made in advance of this review. The existence of order effects suggests that subjects do not use Bayesian principles when integrating evidence relevant to a hypothesis or set of hypotheses. According to Bayes theorem, evidence order should not affect the likelihood assigned to a hypothesis in the light of relevant evidence. As the experimental literature which shall be reviewed in this chapter demonstrates, information order is, very often, a crucial factor in determining a subject's level of belief in a hypothesis. Early studies (Peterson and DuCharme, 1967; Pitz, Downing and Reinhold, 1967) showed that subjects' revision of subjective probabilities falls far short of the prescriptions of Bayes theorem. In general, subjects are not held to revise their beliefs in the light of new evidence in a Bayesian fashion, and Bayes theorem is only referred to in the literature to be reviewed in connection with de-biasing procedures (see Lopes, 1987). Accordingly, Bayesian approaches to the revision of belief will not be discussed in the course of this chapter. The issue of subjects' use of background knowledge about the probabilities of events in their environment will, however, be returned to in later chapters.

3.2. ORDER EFFECTS AND INFORMATION INTEGRATION

3.2.1. The order effects paradigm

The order effects paradigm is very simple and is probably best illustrated using the example of Asch's (1946) impression formation task. In Asch's study subjects were given a list of six adjectives which they were told described an imaginary person. Three of these adjectives were positive - PPP - and three were negative - NNN. Subjects were presented with these adjectives in the order PPPNNN or NNNPPP. Subjects were then asked to write a character description of the target individual. Those subjects who had received the positive adjectives first were found to have formed a more favourable impression of the target individual's personality than had those subjects who received the negative adjectives first. Asch claimed that this finding demonstrated a primacy effect in impression formation and explained his finding thus:
The accounts of the subjects suggest that the first terms set up in most subjects a direction which then exerts a continuous effect on the later terms. When the subject hears the first term, a broad uncrystallised but directed impression is born. The next characteristic comes not as a separate item, but is related to the established direction. Quickly the view formed acquires a certain stability, so that later characteristics are fitted - if conditions permit - to the given direction.

Asch's order effects paradigm was to go on to become to the impression formation literature what Wason's selection task has become to the reasoning literature. In the thirty years between 1960 and 1980 literally hundreds of experiments were run using the order effects paradigm. In fact, it continues to be used to the present day. The domains in which the paradigm has been used have varied but some examples are mock trials and simulated legal evidence (Anderson, 1959; Furnham, 1986; Tetlock, 1983; Walker, Thiabut and Andreoli, 1972); aircraft identification (Adelman, Tolcott and Bresnick, 1993); and auditing (Ashton and Ashton, 1988; 1990). As the above examples suggest, the order effects paradigm has been used in domains other than impression formation. However, the majority of studies have used social materials very similar in nature to those of Asch (e.g. Anderson, 1962; Anderson and Jacobson, 1965; Hendrick and Constantini, 1970; Levin and Schmidt, 1970; Roby, 1967; Stewart, 1965).

One of the problems with the literature on order effects has been the finding of both primacy and recency effects. Although Asch provided evidence for primacy, the earliest demonstration of the importance of information order was that of Lund (1925) who found that recently acquired information was the most important in attitude change. Nisbett and Ross (1980) in a review of the order effects literature argued that primacy effects are the norm in psychological experimentation and are overwhelmingly more probable than are recency effects. This is a gross over-simplification of the literature. In fact, one of the most important studies in the literature on order effects in impression formation is that of Stewart (1965) who found a recency effect using materials very similar to those used by
Asch. The crucial difference in Stewart's study was that subjects were required to rate their impression of the target individual upon receipt of each personality adjective. He found that those subjects who received the positive adjectives in the later stages of the task had a significantly more positive impression of the target individual than had those subjects who received the positive adjectives in the early stages of the task. In a recent paper Hogarth and Einhorn (1992) reviewed the results of 60 papers on order effects and found that 35 of them contained experiments demonstrating the importance of recently acquired information. Of these 35 recency effects 20 involved asking subjects to respond after receipt of each piece of information. Of the remaining 25 studies which demonstrated primacy, only two involved asking subjects to respond after each piece of information.

The implications of this survey are twofold. Firstly, Nisbett and Ross are in error when they claim that primacy effects are more likely. It is clear from Hogarth and Einhorn's survey that both primacy and recency are common findings in the literature. Secondly, requiring subjects to respond as they receive each piece of information is a reliable source of recency effects. A crude (but reasonably accurate, and very common) reading of the effect of the continuous responding manipulation is that it works by causing subjects to attend to each of the adjectives in the series (i.e. by forcing them to treat each adjective as being equally important). The importance of this effect is accentuated further when one examines the fifteen studies which have produced recency effects in the absence of continuous responding. Two of these may be said to have produced recency due to attention manipulations (Anderson, 1968; Hendrick and Constantini, 1970) as subjects were required to pronounce each adjective as they received it. A further four of these studies could be said to rely on a short term memory effect for their results as they involved stimulus materials which subjects may have had difficulty in remembering. For example, Anderson (1967) required subjects to provide an estimate of six weights; Anderson and Jacobson (1968) asked subjects to provide an estimate of three weights; Parducci, Thaler and Anderson (1968) used sets of noise and asked subjects to estimate the average; and finally, Weiss and Anderson (1969) asked subjects to estimate the average length of a set of lines. A reasonable explanation for the results of all of these experiments...
is that recency was found because subjects had difficulty in remembering early occurring stimuli.

The forgetting of early information is also implicated as a factor in at least three of the remaining nine instances where recency in the absence of continuous responding was discovered. Crano (1977) used two 600 word passages in his study of opinion change, whilst Luchins (1957) also used complex stimulus materials. In both of these cases it is easy to imagine how recently presented complex information will seem to be more important than early presented information. Lichtenstein and Srull (1987) found their order effects in the context of a study on memory and judgement. They examined the effect of information order under three conditions. The first two of these - where subjects were told that they would have to make a judgement based on sentences describing a target individual’s behaviour or that they would be asked to recall as many of these sentences as possible - both produced a primacy effect. In the third condition where subjects were told simply to comprehend the information and to check its coherency and grammaticality, a recency effect was discovered. Once again it is likely that recently acquired information seemed to influence the judgement of these subjects due to a very simple memory effect.

It is clear from this review of recency effects in the literature that the vast majority of such effects have some kind of short term memory explanation. Reading the literature one is inclined to conclude that recently acquired information is important only if some experimental manipulation is performed which causes subjects to attend to that recently acquired information. In most other cases, it is early occurring information which is attended to. The question begged by this conclusion is how exactly primacy effects have been explained.

3.2.2. Three Explanations for Primacy Effects

It will be remembered from the discussion of Asch's primacy effect that his explanation for the phenomenon which he had discovered was essentially a gestalt explanation. Asch’s account was the first explanation put forward in the literature for the
existence of primacy effects. Asch argued that later occurring information was selectively processed in terms of the connotative implications of early occurring adjectives. In other words, the meaning of these later occurring adjectives was dependent on the context provided for them by the early occurring adjectives. For fifteen years this interpretation remained unchallenged in the literature. With the exception of one study by Bruner, Shapiro and Tagiuri (1958) which appeared to show that impressions inferred from lists of words could be reliably predicted from an arithmetic averaging of the independent meanings, Asch's own explanation for his primacy effect was the one which was generally accepted.

The first systematic attempt to re-examine Asch's interpretation of his primacy effect was published in 1961 by Anderson and Barrios. These authors reported a study where each subject was presented with over sixty sets of six adjectives. As with the Asch study, these sets were comprised of six adjectives, three of which were positive and three of which were negative. The crucial manipulation was the extent to which the sets of adjectives differed in valence. The valence of an adjective may be defined as its positivity or negativity. Therefore, a set of adjectives which differ sharply in valence might go from extremely positive to extremely negative whereas a set which differs less sharply might go from moderately positive to moderately negative. Half of the adjective sets abruptly changed in valence whilst the other half changed gradually. Once they had been exposed to each set of adjectives subjects were asked to rate their impression of the target individual on a scale which went from -4 (highly unfavourable) to +4 (highly favourable). The results were primacy for both conditions. It seemed to make no difference whether the adjective set contained an abrupt or a gradual change.

Anderson (1981) has characterised the results of this experiment as an initial test of two types of theory of primacy effects. The first of these are interactive theories which suggest that the personality adjectives interact in some way. There are two variants of this position. Firstly there is Asch's "change of meaning" hypothesis described in the introduction to this section which, as has already been mentioned, was the first explanation put forward for primacy. The second of the interactive theories (and, therefore, the second
possible explanation for the existence of primacy effects) is the discounting hypothesis which suggests that the later adjectives are discounted or ignored by the subject as they contradict the impression formed from the early occurring adjectives. Both of the theories involve some interaction between the personality adjectives received by the subject. These theories are in sharp contrast to the attention decrement hypothesis which simply states that the importance of early occurring information is due to a decrease in attention across the series. This, the third possible explanation for primacy effects, is different from either interactive theory in that the attention decrement account does not rely on any relation among the adjectives.

Anderson (1981) has claimed that the results of the Anderson and Barrios study favour the attention decrement account of the primacy effect. This is because of the absence of an effect for type of change in the valence of the personality adjectives. He argues that the interactive hypotheses would both predict a greater primacy effect in the case where the change in valence is abrupt on the grounds that it would make any inconsistencies more salient. The reasoning behind this argument would seem to be suspect however. Whilst it is clear that the Anderson and Barrios study does discriminate between the inconsistency discounting and attention decrement accounts of primacy effects, it is not clear how these results affect the status of the change of meaning hypothesis.

The studies which have manipulated subjects’ attention, discussed in the previous section, have also been claimed as support for the attention decrement hypothesis. For example Anderson and Hubert (1963) read subjects a sequence of personality adjectives and asked them to rate the likeableness of the target individual. However, they also told subjects in certain conditions that there would be a casual recall test of the adjectives after rating likeableness. The results were very clear. Those subjects in the recall conditions failed to demonstrate primacy in their likeableness ratings. In some cases recency was found. This finding has been replicated by Riskey (1979). In a similar vein Stewart (1965) also attempted to arbitrate between the attention decrement and interactive hypotheses. He required subjects in some of his experimental conditions to rate their impression of the
target individual upon receipt of each adjective. Those subjects in the standard conditions, where a response was required only upon receipt of the entire sequence of adjectives, demonstrated the usual primacy effect. Those subjects in the cumulative responding conditions, on the other hand, demonstrated recency.

What both of these studies have in common is their attempt to spread subjects' attention across all of the adjectives in the set. That they both resulted in a failure to find primacy effects suggests that attention decrement is in some way responsible for Asch's original primacy effect. Once again however, it is unsafe to claim that either of these experiments discount Asch's change of meaning hypothesis. Different criticisms may be levelled at each experiment on this score. The Anderson and Hubert study finds no evidence for primacy and some slight evidence for recency under memory set instructions. It is not clear however how these memory set instructions changed the nature of the task. For instance there is evidence from Lichenstein and Srull (1987) that processing objectives (i.e. memory or judgement) affect responding on the impression formation task. It is not clear what subjects' primary processing objective was in the memory set conditions of the Anderson and Hubert study. The presence of the memory instructions not only may have equalised subjects' attention across all of the stimuli contained in the adjective sets, it may also have fundamentally altered the integration process.

The Stewart (1965) study produced clear evidence of recency. On this basis it has regularly been cited by proponents of the "information integration" approach to belief revision as clear evidence that Asch's change of meaning explanation for his primacy effect is in error. This criticism only holds however if it is accepted that the belief revision process is irreversible. There is no doubt that Stewart's recency effect is caused by the manipulation of subjects' attention. What is not so clear is whether subjects are re-interpreting the meaning of earlier adjectives in the light of later adjectives. As has been previously discussed, there exist many demonstrations in the literature of recency effects which rely on attentional manipulations. However, it is not clear to what extent the re-interpretation of earlier evidence may interact with these attention manipulations in
producing these recency effects. The last experiment in this thesis will attempt to shed some light on the issue of the re-interpretation of earlier evidence.

One final set of studies deserves consideration in this section of the chapter. These are the studies which, after the work of Anderson and Barrios (1961), sought to examine the effect of manipulating the inconsistency between the adjectives used in the task. The first study to be discussed is that of Anderson (1965) who varied the amount of inconsistency present between the personality adjectives he included in his sequences. Anderson used two sets of six adjectives. These sets were composed of either three highly favourable adjectives and three moderately favourable adjectives (HM+) or three highly unfavourable adjectives and three moderately unfavourable adjectives (LM-). Half of the subjects received the adjectives in one order, whilst the remaining subjects received them in the other order. With both sets of adjectives a recency effect was obtained. Initially this result was taken to support the discounting hypothesis. If the primacy effect with sets of highly favourable and highly unfavourable adjectives is due to discounting then it should vanish when only positive or negative adjectives are used (because the discounting of later occurring information is unnecessary when the all of the adjectives are positive or negative). On the other hand, the attention decrement hypothesis predicts that primacy will be obtained with any set of adjectives.

Hendrick and Constantini (1970) speculated that Anderson's result was due not to the reduction of the inconsistency between the adjectives but instead to his requirement that subjects read each adjective aloud as it was exposed. They claimed that this part of Anderson's procedure was likely to have caused subjects to attend equally to each adjective in the set thus reducing attention decrement. In order to test this intuition Hendrick and Constantini carried out two experiments. In the first of these experiments the consistency between the adjectives present in the sets used was systematically varied. This was achieved by means of a pre-test where subjects were asked to estimate the probability that a person possessing a set of three favourable attributes could also produce a set of three unfavourable attributes. In this way Hendrick and Constantini were able to construct sets of adjectives that contained sets of positive and negative traits which were perceived to be
of either high or low relatedness. Once again, the attention decrement hypothesis predicts primacy effects in both high and low relatedness conditions. The discounting hypothesis predicts more primacy in the low relatedness condition. As predicted by attention decrement, primacy was found in both conditions.

In this first experiment by Hendrick and Constantini subjects were not required to read each adjective as it was revealed. In a second experiment using exactly the same procedure, half of the subjects read the words as they were presented to them whilst the other half did not. The results were clear. In the conditions where subjects pronounced the traits Anderson's recency effect was replicated. On the other hand, in the conditions where subjects did not have to read the adjectives, the primacy effect from Hendrick and Constantini's first experiment was replicated.

Whilst the experiments just described clearly discriminate between the attention decrement and discounting hypotheses, once again it is not clear how they affect the status of the change of meaning hypothesis. As with the attempts to rule out all forms of interactive account which were described earlier, the above experiments do not rule out a change of meaning account. This is because these experiments were not designed with the change of meaning hypothesis in mind. Regardless of the size of the difference in polarity between the two sets of adjectives, the change of meaning account would still predict a primacy effect which is exactly what Hendrick and Constantini found in their first experiment.

The purpose of the experiments described above was to rule out interactive accounts of how the order in which information is received determines the effect of that information on personality impressions. It is clear that much of the evidence produced argues against a discounting explanation of subjects' behaviour although this may be a product of the impression formation task itself. It is hard to imagine that cases do not exist where context will cause subjects to discount some of the information which they receive. What is not so clear is the status of Asch's change of meaning hypothesis in the light of this evidence. It may be that the problem with Asch's account is its lack of specificity in
terms of cognitive process. This lack of specificity means that it is very hard to completely rule out the change of meaning account. However, evidence will be discussed in a later section of this chapter, which suggests how Asch's original insight might be made more specific.

Before discussing that work however, it will be necessary to do two things. Firstly, the initial models induced from the extensive work on order effects in belief revision must be discussed. These models are very different to the model which will be sketched at the end of this chapter. Secondly, more recent models which have their roots in these initial models must also be discussed. Although these more recent models are a marked advance on the earlier models it will be argued that they are inadequate. What both sets of models have in common is that they attempt to capture the nature of the belief revision process by the application of mathematical operations to the pieces of evidence which subjects receive. It will be argued that whilst they may very well capture the existing data, there are several a priori reasons for doubting their ultimate usefulness.

3.3. MODELS OF BELIEF REVISION AS INFORMATION INTEGRATION

3.3.1. Cognitive Algebraic Models

From the previous section it is apparent that the major worker in the area of order effects and information integration is Norman Anderson. Not only did Anderson and his co-workers produce most of the empirical evidence regarding the order effects paradigm, they also did most of the early work attempting to model the process. This work is underpinned by the assumptions of cognitive algebra, an approach to cognition which although not as influential as it once was, still continues to inspire some theoretical and experimental work (see Schlottmann and Anderson, 1993 for an example of information integration theory applied to children's perception of causality).

The cognitive algebraic approach to cognition is based on the fact that subjects, when attempting to integrate several pieces of information, very often seem to be
following algebraic rules (Anderson, 1981). At various times they seem to be adding, subtracting or multiplying when attempting to put the information together. The study and description of these kinds of algebraic phenomena rest on several concepts of which cognitive algebra is but one. Shanteau and Nagy (1984) have illustrated the basic elements of the Information Integration approach to cognition. Their analysis is represented in Figure 3.1 below. As may be seen from Figure 3.1, there are three main components of the information integration process. The first of these is the evaluation of the stimuli. Once evaluated the stimulus comes to have a subjective value for the subject. This subjective value is derived

![Diagram](image)

**Fig 3.1: Diagrammatic representation of the Shanteau and Nagy conceptualisation of Information Integration.**

from two parameters. The first of these is the subjective stimulus scale value \( s \) for the explicit stimulus. This may be regarded as the evidential strength of any piece of evidence in any given context. For example, in an impression formation task the adjective "humorous" may be perceived as highly favourable when an impression of a target individual's likeableness is being formed. On the other hand, if the impression being formed was of the target individual's suitability for a job as an undertaker, the effect might not be quite the same.

In one sense, therefore, the notion of subjective scale value is capable of dealing with the context in which a piece of evidence is encountered. However, such a notion of subjective value only accounts for the context in which the judgement is made. It will be remembered that the difference between attention decrement and interactive accounts of
order effects in personality impression formation tasks was due to the hypothesised effect of the surrounding adjectives on the evaluation of any individual adjective. In order to account for the evidential context of evaluation proponents of the cognitive algebraic approach to cognition have proposed a second parameter. This second parameter is the weight (w) of any piece of information. Although Shanteau and Nagy (1984) claim that the weight value "can generally be thought of as the importance or relevance of" an evidential item (pg. 50), in practice the weight parameter is never invoked in this way. Once again returning to the literature on order effects in impression formation, it is apparent that the weight function is normally invoked to account for the decrement in attention which researchers like Anderson and Shanteau hold to be the cause of primacy effects. Such an argument claims that the later an adjective occurs in the set of adjectives, the less is the weight assigned to that adjective. As the subjective value of each adjective is adjusted by its weight, later occurring adjectives have a smaller impact on the final rating than do early occurring adjectives.

The second component of the information integration approach is the integration function itself. This second component determines how the subjective, weighted values for the evidence are combined in order to produce a judgement. There have been many different types of models suggested (see Anderson, 1981 for a review) but most discussion has centred around additive and averaging models. As their names suggest both types of model are based on the very simple intuition that subjects combine evidence by the application of very simple mathematical operations. A very simple additive model is represented in Equation 3.1.

\[ r = s_1 + s_2 \]  
Eq.3.1

This is a model of the case where a subject receives two pieces of evidence \((s_1\) and \(s_2\)) and derives her response by adding the subjective values for the pieces of evidence. In general this model is too simple too account for subjects' behaviour. For this reason, models such as that described in Equation 3.2 were devised.
Because the weight assigned to a piece of evidence decreases the later the piece of
evidence occurs in the sequence, this kind of model captures the primacy effect found on
the impression formation task. Shanteau (1970; 1972) is one author who initially favoured
additive models of information integration. Indeed, whilst the goal of the research carried
out by proponents of information integration theory was simply to discriminate between
attention decrement and interactive accounts of Asch's primacy effect, additive models
were preferred to all other types of model.

Unfortunately however, the additive model is not the only algebraic model which
predicts a primacy effect in impression formation. Consider the model expressed in
Equation 3.3.

\[ r = \frac{s_1 + s_2}{2} \]  

Eq.3.3

This is the simplest form of averaging model to describe how subjects integrate two pieces
of evidence. Essentially it states that subjects arrive at a judgment by adding the
subjective scale values for the evidence and then taking an average. In other words,
subjects' impression of a target individual or belief in a hypothesis is equal to the average
strength of the evidence which they have received. The weighted average version of this
model, given in Equation 3.4 makes exactly the same primacy predictions for the
impression formation task as does the weighted adding model.

\[ r = \sum_{i=0}^{N} w_i s_i / \sum_{i=0}^{N} w_i \]  

Eq.3.4

The weighted average model is very easy to explain. Imagine the subject asked to
rate the likeability of a target individual. She is presented with a series of three adjectives
which describe that individual and have subjective values of 0.1, 0.4, 0.2 respectively (where a
low subjective value corresponds to a negative adjective and a high value corresponds to a
positive adjective). Each of these scale values will be assigned a weight by the subject (amongst information integration theorists the value of this weight is generally held to depend on where a particular adjective comes in the sequence - early occurring adjectives receive heavier weighting). For the purposes of this example, suppose that the weights assigned are .6, .4, and .2 respectively. The weighted averaging model of information integration predicts that the subject's response (see Eq. 3.4a) is simply the sum of the weighted scale values divided by the sum of the weights.

\[ R = \frac{1(.6) + .4(.4) + .2(.2)}{.6 + .4 + .2} \]  

Eq. 3.4a

That the averaging and additive models make the same predictions in many cases becomes clear if one considers a simple impression formation task where the subject is told that the target individual has been described by one acquaintance as being humorous (a highly favourable trait) and as being cold (a highly unfavourable trait) by another. With both averaging and additive models primacy will occur because the first trait (whether it is favourable or unfavourable) will be more heavily weighted than the second. It does not matter that the mathematical operation embodied in each model is different - in both cases the greater weighting of the first adjective will result in primacy.

How then are the averaging and additive accounts to be separated? Although they may seem to be very similar, averaging and adding models do, in certain situations, make differing predictions. This is due to the fact that averaging models possess two types of weighting. The first of these is the explicit weighting, described above. The second is the weighting implicit in all averages - every value which goes into the production of the average weights every other value. Once it was felt that interactive accounts of impression formation had been ruled out (although as has already been pointed out, this is not the case) testing the differing predictions made by averaging and additive models became the central research question in the information integration literature. Several elegant tests were devised, the first of which was reported by Anderson (1965). He asked subjects to rate the likeableness of target persons described by sets of either two or four adjectives. The two adjective sets were composed of either two highly favourable adjectives (HH) or
two highly unfavourable adjectives (LL). The four adjective sets were comprised of two highly favourable adjectives and two moderately favourable adjectives (HHM+M+) or two highly unfavourable adjectives and two moderately unfavourable adjectives (LLM-M-).

The averaging and additive models diverge in their predictions about the above stimulus materials. If an additive function is in operation then the evaluation of the target person given two highly favourable adjectives should be higher than that given highly favourable and moderately favourable adjectives. Similarly, the evaluation of the target individual should be lower given two highly unfavourable adjectives than it would be given two highly unfavourable adjectives and two moderately unfavourable adjectives. If an averaging function is in operation however, these predictions are reversed. The average of two highly and moderately favourable or unfavourable adjectives should always be less extreme than the average of two highly favourable or unfavourable adjectives. With both favourable and unfavourable adjectives Anderson's results supported the averaging model. That is \( R_{HH} > R_{HHM+M+} \) and \( R_{LL} < R_{LLM-M-} \). This result has been replicated by Hendrick (1968) and by Leon, Odeon, and Anderson (1973).

Similar results have been reported by Lampel and Anderson (1968) who used a procedure asking female college students to rate the dateability of males described by either a pair of personality trait adjectives and a photograph or a photograph alone.

As many of the criticisms which may be levelled at the information integration approach to impression formation, also apply to later models of belief revision, they will not be discussed here. Instead, the criticism which is unique to the impression formation approach will be briefly discussed. This criticism is based on the fact that information integration theory is enormously ambitious and very powerful but possibly built on unsound foundations. The approach is based on three assumptions. The first of these is that subjects' ratings of likeability, dateability etc. are made on an interval scale. This assumption may or may not be valid. The second assumption is that subjects use some kind of mathematical function to integrate the subjective scale values of pieces of
evidence. The third is that statistical techniques such as the Anova may be used to uncover both these subjective values and the weights which are assigned to them. In a sense, these assumptions are reliant on each other for their validity. Anderson (1981) points to the vast number of experimental results which corroborate the interval scale assumption. Yet these results are themselves the result of fully factorial designs where subjects' responses are analysed using Analysis of Variance.

Furthermore, the estimation of weight and scale values is dependent on the assumption that subjects are using an averaging rule. Shanteau and Nagy (1984) have admitted that the integration rule and the estimation of weight and scale values are dependent on each other. If this is the case, then both are dependent on the assumption that subjects are responding on an interval scale. This assumption in turn underlies the extensive use of the Anova. The problem with the approach is neatly summed up by Lovie (1984):

*One can, therefore, view Anderson's system of functional measurement as a kind of elaborate balancing act where scale type (i.e. interval), substantive concepts (i.e. integration and averaging theory) and analysis (i.e. analysis of variance) simultaneously support and justify each other. Failure of any of the parts could, therefore, bring down the whole edifice, but, because of the interdependency of the structure it would, paradoxically, be difficult to determine which part had failed.* (pg. 94)

### 3.3.2. Procedural Models

Although the cognitive algebraic approach to cognition was very productive in the 1960's and 70's, its star has, to a large extent, waned. However, rather than being discarded due to evidence which suggested that the approach was ultimately unworkable, it has simply been replaced by a new generation of procedural models (Lopes, 1987; Hogarth and Einhorn, 1992). That the information integration approach has never been demonstrated to be inadequate is probably due to what Lakatos (1970) calls *inoculation.*
Although peripheral theoretical details may have changed (e.g. the substitution of averaging models for additive ones), the core concepts have remained the same. As the criticism at the end of the last sub-section implies, the manner in which these concepts relate to one another suggests that it will be very difficult to provide truly discriminating evidence.

Instead emphasis has simply shifted within the literature. The new generation of procedural models, as their name suggests, seek to model the process of belief revision. Nowhere is this better exemplified than in the work of Lopes (1987). Taking as her starting point the fact that subjects seem to use an averaging procedure in judgement tasks, she has attempted to elucidate other aspects of the judgement process which may lead to judgmental error. In so doing she seeks to induce a model of what people do when they make judgements rather than of the data that they produce (pg. 169). Essentially her process model claims that averaging is produced because people integrate new information with old composite judgements by adjusting the old value so that the new composite lies somewhere between the old composite and the value of the new information. As she points out, this process is qualitatively similar to averaging but does not presuppose that subjects ever "compute" an average in the algebraic sense of the term.

Lopes' model is a very simple one and is composed of just a few stages. Firstly, it is proposed that subjects scan the evidence. Once this has been done an evidential item is chosen as the "anchor point". Typically, the anchor will be chosen based on its relative importance (e.g. its diagnosticity). Once an anchor has been chosen, it will be evaluated relative to the scale of judgement. In many cases this will produce an initial judgement. Subsequent to this, remaining items are integrated in the "averaging" manner described above. The remaining items are integrated in order of importance and the process stops when there are no important items left unaccounted for.

This kind of model is certainly an advance, in psychological terms, on the earlier cognitive algebraic models. Most important has been the notion of an anchor. Its centrality is evidenced by the psychological importance ascribed to it in a more recent, and more
comprehensive, model of the belief revision process. This model will now be discussed in some detail.

The belief-adjustment model of Hogarth and Einhorn (1992) is the most ambitious of current approaches purporting to model order effects in evidence integration. Their model assumes

...that people handle belief-updating tasks by a general, sequential anchoring-and-adjustment process in which current opinion, or the anchor, is adjusted by the impact of succeeding pieces of evidence.

In its most basic form their model may be expressed

\[ S_k = S_{k-1} + w_k [s(x_k) - R] \]

Eq. 3.5

where \( S_k \) equals the degree of belief after \( k \) pieces of evidence \((0 \leq S_k \leq 1)\), \( S_{k-1} \) corresponds to the anchor or prior opinion (initial strength of belief is \( S_0 \)), \( s(x_k) \) equals the subjective evaluation of the \( k \)th piece of evidence, \( R \) equals the reference point or background against which the \( k \)th piece of evidence is evaluated and \( w_k \) equals the adjustment weight for the \( k \)th piece of evidence \((0 \leq W_k \leq 1)\).

The model expressed in Equation 3.5 is very simple. It is useless however without some definition of its components. Unfortunately, therefore, things are not as simple as Equation 3.5 may seem to suggest. In line with Lopes' emphasis on what subjects actually do when they update their beliefs, Hogarth and Einhorn propose three different sub-processes which constrain the operation of the processes described by the simple form of the model. These are (1) how evidence is encoded, (2) how evidence is processed, and (3) how the adjustment is accomplished. Each of these sub-processes and its effect on the operation of the simple model will be considered individually. As this discussion will become quite technical and very involved, the weaknesses inherent in Hogarth and
Einhorn’s characterisation of each of these sub-processes will be presented as the sub-process is described.

3.3.2.1. How the evidence is encoded.

The first sub-process which Hogarth and Einhorn consider is the encoding of evidence. By this they mean the process by which the evidence is related to the hypothesis. They define two different types of relationship in terms of whether the evidence is encoded relative to constant or variable reference points. Furthermore, they claim that any belief revision task may be categorised as involving evaluation or estimation. In evaluation tasks a constant reference point is used. Evidence is encoded as being positive or negative relative to a theory or hypothesis. For example, consider a research scientist trying to decide whether a given hypothesis is valid where her belief about the hypothesis ranges along a continuum from "false" (=0) to "true" (=1). In this situation positive evidence always increases belief and negative evidence always decreases it. In other words, regardless of the scientist’s current level of belief in the hypothesis, supporting evidence always increases belief in the hypothesis whereas disconfirming evidence always decreases it. Thus, evidence is always bipolar (-1 ≤ s(x_k) ≤ +1), and prior belief always equal to zero (R = 0). Equation 3.5 can be rewritten for evaluation tasks as

$$S_i = S_{i-1} + w_s(x_i)$$

Eq.3.6

In estimation tasks, on the other hand, evidence is encoded relative to variable reference points. These tasks involve some kind of “moving average” and each new piece of evidence is encoded relative to current opinion. For example, the typical impression formation task, which involves an impression of “likeableness”, is a good example of this kind of task. Current opinion may be thought of as being expressed on a continuum from "dislike" (=0) to "like a lot" (=1). Any personality adjective given to a subject will increase how much they like the target individual only if that adjective is more positive than the subject’s current opinion of the target. Therefore, evidence is encoded as being positive or negative relative to the current level of belief rather than relative to the hypothesis under
consideration. Contrary to evidence encoding on evaluation tasks, evidence is unipolar \((0 \leq s(x_k) \leq 1)\) and \(R = S_{k-1}\). Accordingly, Equation 3.5 may be rewritten as

\[
S_k = (1 - w_k)S_{k-1} + w_k s(x)
\]

Eq. 3.7

As Hogarth and Einhorn point out, Equation 3.5 corresponds to an adding formulation for evidence integration, whereas Equation 3.7 corresponds to an averaging formulation. The advance which this approach makes on previous approaches is that it has provided a rationale for the differential use of adding and averaging formulations, both of which may be seen as by-products of what seems, initially at least, a psychologically real distinction.

It is the issue of what is psychologically real that is of most trouble to this aspect of the Hogarth and Einhorn model. Whilst a distinction between the encoding of evidence in terms of either a fixed or moving reference point seems plausible and attractive at first glance, the question remains as to whether it is either theoretically or experimentally tenable. Evaluation tasks, all of which involve fixed reference points, are characterised by one or more of the following:

- evidence is coded as positive or negative relative to the hypothesis under consideration. For example, evidence relevant to the truth of a hypothesis will either confirm or disconfirm that hypothesis.

- evidence is conceptually measured on a bipolar scale \((-1 \leq s(x_k) \leq 1)\), and then transformed onto a unipolar scale \((0 \leq S_k \leq 1)\). In other words, the evidence which a subject receives may be positive or negative but is used to adjust belief on a scale which runs from "false" (=0) to true (=1).

- evidence is marked for one hypothesis or another (Lopes 1985) which facilitates a directional interpretation of the information. If evidence is marked as positive or negative this makes it easier to understand, and use, in terms of confirmation and disconfirmation.
• evidence requires explicit interpretation as to whether it is for or against a hypothesis. Once again, Hogarth and Einhorn claim that this facilitates interpretation of evidence in terms of confirmation or disconfirmation. They give the example of a juror having to decide whether certain evidence supports a verdict of guilt or innocence. Because the hypothesis is bipolar in nature the evidence must be explicitly interpreted as to which verdict it supports.

Estimation tasks, on the other hand, which involve moving reference points, are characterised by one or more of the features below:

• evidence is interpreted relative to current opinion - subjects are sensitive to the difference between the location of the current anchor and the level of belief associated with the current evidence

• evidence is transformed from one unipolar scale \((0 \leq s(x_k) \leq 1)\) onto another \((0 \leq S_k \leq 1)\). That is, evidence is never negative as it is always interpreted in terms of its relationship to existing evidence and is used to adjust belief along a continuum which ranges from, for example, "dislike" (=0) to "like a lot" (=1)

• evidence integration is associated with questions of "how much" rather than "true-false". Indeed, continuous scales may elicit estimation whereas dichotomies may elicit evaluation. This may be thought of in terms of the earlier distinction between deciding whether a hypothesis is true or false and deciding how much one likes a target individual.

• evidence requires very little interpretation as to whether it is for or against the hypothesis under consideration. This facilitates a comparison of the new evidence with already existing levels of belief. For example, imagine being asked to decide whether you like or dislike some individual. If you already possess some opinion, being told
that five out of six of that person's acquaintances dislike him requires very little interpretation and is easily compared to one's existing levels of belief.

With such a list of defining characteristics the distinction between evaluation and estimation tasks might seem to be clear. Closer inspection reveals that this is not so. Several points need clarification. Firstly, it is not clear that the characteristics of evaluation or estimation tasks are mutually exclusive. Various tasks might be designed which confound certain characteristics of both kinds of tasks. For example, an alternatives task (true/false) which required subjects to integrate numerical information, or an estimation task which used pieces of evidence clearly marked for one hypothesis or another. An example of the former case might be asking subjects to decide whether a given hypothesis was true or false based on Bayesian evidence (i.e. base rates and likelihood ratios). Bayesian evidence is unipolar yet deciding whether a hypothesis is true or false is an evaluation task which requires that evidence be interpreted on a bipolar scale. An example of the latter case (an estimation task where evidence is clearly marked for one hypothesis or the other) might be asking subjects the degree to which they believe that a target individual, who possesses some characteristics, comes from one of two occupational groups (as in Tversky and Kahneman's, 1974, demonstration of the representativeness heuristic). Imagine that subjects receive additional sequential evidence about the occupation of specific individuals who possess the same characteristics on this task. The task is one of estimation (i.e. subjects must express their degree of belief) yet the evidence is clearly marked for one or other of the hypotheses.

Two counter-arguments are possible. Firstly, it may be possible to argue that there are a range of special cases where characteristics of estimation and evaluation tasks are to be found in the same task, but that usually tasks are structured as Hogarth and Einhorn claim they are. The problem with this kind of counter-argument is that it diminishes the power of the model. The second counter-argument that is possible is that although these features tend to cluster together in evaluation and estimation tasks, there is really one defining feature of each type of task around which the other features tend to cluster. The problem with this type of counter-argument is that it is almost always possible to imagine a
version of a specific task where one of the features is reversed. For example, if it might be claimed that the defining difference between evaluation and estimation tasks was whether evidence was used to evaluate a phenomenon's category membership (Is this person guilty or innocent? or Is this hypothesis true or false?) or the degree to which a particular state of affairs pertains (How much do I like this person? or How likely is it that it will snow tomorrow?). However, it is always possible to ask subjects to make a categorical decision about whether they like someone or to estimate how likely it is that a hypothesis is true or false.

A second problem is the distinction made between unipolar and bipolar scales. Allied to this distinction is whether evidence is encoded relative to the hypothesis under consideration (bipolar scale), or to evidence already encoded. Mynatt, Doherty, and Dragan (1993) have demonstrated that in the choice of evidence on a diagnostic reasoning task subjects are constrained by the hypothesis to which subjects are attending. The hypothesis attended to can be manipulated by the direction of the first piece of evidence. Current theorising on hypothetical thinking (e.g. Evans and Over, 1996b) emphasises the fact that subjects represent only one alternative at a time. Although the Hogarth and Einhorn model can handle the case where subjects are trying to determine whether or not one alternative is likely to be the case, it does not give an account of what happens when subjects are trying to decide between two alternatives. For example, what happens when subjects are trying to decide between two alternatives where the evidential set consists of evidence explicitly marked for one or other alternative? Do subjects flip from one scale to another? Or does the task become one of estimation?

### 3.3.2.2 How the evidence is processed

The second sub-process which Hogarth and Einhorn consider is the processing of evidence. Once again they draw a distinction between two forms of processing. The key issue here is whether the evidence is processed step by step (SbS), or at the end of a series (EoS) of pieces of evidence. These two distinct processing strategies map onto the experimental manipulation discussed in the second section of this chapter which seems so
crucial to whether a primacy or recency effect is found. In the SbS process belief is adjusted incrementally by each piece of evidence whereas in the EoS process the initial anchor (formed from the first one or two pieces of evidence) is adjusted by the net effect of the remaining evidence. The EoS anchoring and adjustment strategy is represented as follows

\[ S_t = S_0 + w_t[s(x_1, ..., x_k) - R] \]  

Eq. 3.8

It is slightly unclear as to what \( s(x_1, ..., x_k) \) represents although Hogarth and Einhorn say it is "...some function, possibly weighted average, of the individual subjective evaluations of the items of evidence that follow the anchor."

It is claimed that choice of processing strategy is determined by the demands of the task. Thus an SbS strategy is used when

- the task demands an SbS response mode
- the relative complexity and/or length of the series is small

On the other hand subjects choose EoS strategies when

- the task allows for an EoS response mode (it is important to realise however that it is always possible to use an SbS strategy, even when response mode is EoS. SbS will be used if complexity and length of series factors are appropriate)
- the evidence consists of simple items in short series.

This is the most non-contentious part of Hogarth and Einhorn's model. It seems to capture the distinction between sequential responding and responding at the end of the series which has been found so important in determining the source of order effect. However, this distinction does seem to make parts of the earlier distinction redundant. Specifically, is it
possible that a task which requires SbS responding can ever be one of evaluation? It will be remembered that in evaluation tasks R (the background against which a piece of evidence is evaluated) is always equal to 0. Although this may be possible for the first piece of evidence which the subject receives in an SbS task, it is not possible thereafter. Once the subject has received some evidence R can no longer be equal to 0. As with the distinction between evaluation and estimation tasks, it is likely that the distinction between SbS and EoS responding is not as clear as it might initially seem.

### 3.3.2.3 How the adjustment is accomplished

The third sub-process in the updating of beliefs considered by Hogarth and Einhorn is the process of adjusting belief in the face of evidence. This adjustment is common to both evaluation and estimation tasks regardless of whether it occurs step by step or at the end of a series. Hogarth and Einhorn's model of how this is accomplished stems directly from what they call a "contrast assumption". This contrast assumption is embodied in their claim that the adjustment weight should depend on whether the new piece of evidence is more or less positive than background knowledge \((s(x_k) - R)\) in Equation 3.5) and the level of the anchor \(S_{k-1}\). Accordingly, they propose two separate equations for the adjustment weight.

The first of these is for cases where \(s(x_k) \leq R\):

\[
w_i = \alpha(S_{i-1})
\]

Eq. 3.9

Substituting Equation 3.9 into Equation 3.5, we now have

\[
S_i = S_{i-1} + \alpha(S_{i-1})(s(x_i) - R)
\]

Eq. 3.9a

The second case is for when \(s(x_k) \geq R\):

\[
w_i = \beta(1 - S_{i-1})
\]

Eq. 3.10
Once again, substituting in Equation 3.5 we get

\[ S_k = S_{k-1} + \beta(1 - S_{k-1})[s(x_k) - R] \quad \text{Eq. 3.10a} \]

From Equations 3.9 and 3.10 it can be seen that \( w_k \) is proportional to the anchor when the subjective value of the new evidence is less than \( R \) and inversely proportional to the anchor when the value of new evidence is greater than \( R \). This means that when the anchor is great and the new evidence is weaker than \( R \), this new evidence will have the effect of a large downwards revision of belief. When the anchor is low and the new evidence is weaker than the background belief this will result in very little downwards revision (of course the amount in both cases is related to how much weaker than background belief the new evidence is). This of course makes sense. If one does not hold a position very strongly then negative evidence will not effect that weak prior level of belief as much as it will a strong prior level of belief.

Conversely, when the subjective value of the new evidence is greater than \( R \) and the anchor is great, there will be little upwards revision. On the other hand, when the subjective value of the new evidence is greater than \( R \) but the anchor is low, there will be large upwards revision of belief. Once again, this makes perfect sense. If one holds a belief strongly then evidence in favour of that belief cannot cause much upwards revision. Strong positive evidence will cause large upwards revision, on the other hand, when the original belief is weakly held. From the foregoing it is clear that the contrast assumption results in predictions of large revisions of belief when the difference between the subjective value of new evidence is either much greater or much weaker than the strength of background belief. The implications of this assumption will be discussed in the next section.

Finally, it should be pointed out that the constants \( \alpha \) and \( \beta \) refer to subjects' sensitivity to negative and positive evidence respectively, where \( 0 \leq \alpha, \beta \leq 1 \). They represent an attempt by Hogarth and Einhorn to account for individual and contextual factors in sensitivity to new evidence.
Hogarth and Einhorn's model makes very specific predictions concerning order effects. For example, it always predicts recency when subjects are responding in a step by step fashion and $R = S_{k-1}$ (see Hogarth and Einhorn, 1992 for a formal derivation of this prediction). As discussed in the second section of this chapter, SbS responding has always resulted in recency. For end of series responding Hogarth and Einhorn's model always predicts primacy. This prediction is justified by the assumption that as the first piece of evidence in the series will almost always be the anchor it will be weighted more heavily than the aggregate of the remaining evidence by which it is adjusted. Of 27 studies EoS studies which Hogarth and Einhorn review, 19 result in primacy. It is significant that in their assumption that the initial piece of evidence will always serve as the anchor in EoS responding, Hogarth and Einhorn disagree with Lopes who claims that the most important piece of information will be chosen as the anchor.

Many of Hogarth and Einhorn's predictions have never been tested. For example they make predictions concerning complex pieces of evidence for which they can find no match in the experimental literature. However, those experiments which have been run, mostly in applied domains, as tests of the model (e.g. Ashton and Ashton, 1988; Tubbs, Messier, and Knechel, 1990; Asare, 1989; Koch, Pei and Reed, 1989) have all resulted in confirmation of the model's predictions. Initially it would seem that the Hogarth and Einhorn model not only captures most of the existing data, it has also passed several experimental tests. However, there is a major, and perhaps insurmountable, problem with the information integration approach to belief revision. This problem is suffered by both cognitive algebraic and procedural models. The next sub-section will consist of a discussion of these problems.

### 3.3.3. A Problem for Existing Models

Although both Lopes and Einhorn and Hogarth have drawn a distinction between their approaches to the belief revision process and those of workers such as Anderson and Shanteau, both algebraic and process models will be considered together here. This is
because they share the same basic assumption about human cognition. This assumption is best expressed by Anderson (1981), who describes it as:

...an important, implicit assumption that requires explicit discussion. This is the independence assumption that the scale value of each stimulus is constant, independent of what other stimuli it is combined with (pg. 18).

It is this assumption which underlies the entire information integration approach to order effects in belief revision. Accordingly, while Asch claimed that the adjectives which subjects received in his personality impression formation task interacted, the information integration approach claims that the stimuli never interact. This assumption leads directly to what is known as the parallelism analysis. Suppose that two or more stimulus variables are supposed to add together to yield an observed response. If these variables are manipulated in a factorial design then the factorial plot of the data should be parallel.

A good example of a parallelism analysis in operation is contained in a study by Anderson (1962). He asked subjects to rate the likeableness of a target individual in the light of two personality adjectives. Each two adjective set was composed of one adjective from the set: level-headed, unsophisticated, ungrateful; and one adjective from the set: good-natured, bold, humourless. Thus, the experiment had a 3x3 factorial design, with subjects receiving all nine possible pairs of adjectives. The response plots for each subject exhibited remarkable parallelism. Anderson argued that if impression formation was achieved by the interaction of personality adjectives, then these interactions should have manifested themselves in the plot of the data. Although, as was discussed in a previous sub-section, additive models have to a large extent been ruled out, and marked non-parallelism has often been found in information integration experiments (e.g. Lampel and Anderson, 1968), the basic assumption still remains: The manner in which humans integrate information is best described by assuming that stimuli do not interact. Where it seems that they may be interacting (as in the cases where non-parallelism has been observed) this is claimed to be due to the operation of a differential weighting operation. That is, some stimuli are more important than others in certain contexts. It will be
remembered that this was exactly the explanation given for the observed non-parallelism observed in the Lampel and Anderson experiment.

Although the independence assumption is not emphasised by the more recent researchers whose aim is to model the process of belief revision, it is still implicitly contained in their models. Lopes' model, although not formally expressed, accounts for averaging behaviour in terms of difference between the strength of a piece of evidence and the strength of a currently held opinion. Likewise, Hogarth and Einhorn assume that the process may be modelled in terms of the difference between the strength of the new evidential item relative to the strength of background belief. None of the existing models contain either the pragmatic or inferential component which is currently held to be so crucial in models of many aspects of cognition (see Shannon, 1988; Sperber and Wilson, 1986; 1995; Evans and Over, 1996; Hilton, 1995).

There are several a priori reasons for supposing that any model of the belief revision process which does not have an inferential pragmatic component cannot hope to be adequate. Indeed, there is some recent experimental work which explicitly challenges the stimulus independence interpretation of the impression formation literature. There is also older work in the concept literature which is directly relevant, as is some work on representational theories of meaning. All three strands of work, and their implications, will be individually discussed.

3.3.3.1 Inference, domain-specific knowledge structures, and information integration

Recent experimental work by William Wattenmaker (1995) has several implications for current models of information integration. Wattenmaker's main interest is in domain specific knowledge structures and their effect on the learnability of category structures. He proposes the existence of at least two types of categories. The first of these are categories which are linearly separable. That is, these categories may be partitioned on the basis of a weighted, additive combination of component information. If the features of two category members may be weighted and summed so that there is no overlap between
the categories, then these categories may be said to be linearly separable. Conversely, non-linearly separable categories are those where there is overlap once the features have been weighted and summed. Earlier work by Wattenmaker, Dewey, Murphy, and Medin (1986) had shown that the ease with which abstract categories are learned is dependent on how compatible the category is with the knowledge which is activated whilst learning is taking place. For linearly separable categories, any knowledge which provides a basis for the summation of individual features was found to facilitate learning.

The example which Wattenmaker gives is that of the subject who receives the following description: knowledgeable, competent, hard, and composed. In the absence of a context no knowledge is available to suggest a way to integrate these features. If a context such as "scientist" is provided, then the subject can draw on background knowledge to enable her to evaluate and sum the features. Wattenmaker et al found that linearly separable categories were easier to learn when subjects were given themes which provided the basis for summing individual features. On the other hand, Wattenmaker et al found that with non-linearly separable categories which contained correlated features, knowledge that highlighted these correlations made the NLS categories easier to learn. Such knowledge however, makes LS categories more difficult to learn. Wattenmaker et al concluded that the ease with which a category will be learned is determined by the compatibility between the integration strategies suggested by background knowledge and the abstract structure of the category itself.

Wattenmaker (1995) sought to investigate the types of integration strategies induced by knowledge in different domains. He distinguished between social categorisation (the categorisation of people based on traits or behaviour) and object categorisation (the categorisation of concrete entities in the environment). There are good reasons for supposing, in advance, that different knowledge structures underlie object and social categorisation. Wattenmaker identifies three such reasons. As they are relevant, not only to a description of Wattenmaker's work, but also to the issue of what subjects actually do in belief revision tasks, they will be described in some detail.
The first reason for supposing that object and social knowledge have different underlying structures is the ease with which inconsistent social information may be integrated. Lingle, Alton and Medin (1984) analysed social and object categories and concluded that a major difference between the two was the greater flexibility associated with social knowledge. Accordingly, although it is very easy to form hierarchies of object categories which possess such properties as transivity and mutual exclusivity, this is very often impossible with social categories. Consider the academic who resigns her post. This can be interpreted as providing evidence that the academic in question may be categorised as mad, disillusioned, busy, insightful, principled, or even rich! Because of the flexibility of such knowledge, it is easier to integrate inconsistent information in social domains. This is not simply based on intuition. There is evidence (Asch and Zukier, 1984; Kunda, Miller and Claire, 1990) that subjects can resolve inconsistencies in social knowledge with very little difficulty. Both Asch and Zukier and Kunda et al have provided evidence that the main mechanism by which this is achieved is causal reasoning. People infer a reason for the inconsistent information.

On the other hand, inconsistencies in object information are not so easy to resolve. Consider an animal said to have wings, feathers, a beak, and four legs. Although such an animal possesses three very good features for classification as a bird, it is difficult to imagine how the animal could be so classified.

The second reason for supposing that there are domain specific differences in knowledge organisation is related to the idea of flexibility of feature interpretation. This is the amount of overlap which exists amongst social concepts. As it is only possible to sum features when they are perceived to possess some common property which provides a basis for the summation, the flexibility with which social features may be interpreted makes it more likely that such a common property will be found. This flexibility also makes it easier to interpret features in terms of a given category label.

The third reason for supposing differences due to domain in knowledge organisation is the evidence which exists suggesting that social knowledge is represented
differently to object knowledge. It has often been proposed (e.g. Homa, Sterling and Trepel, 1981; Smith and Medin, 1981) that concepts may be represented by a combination of exemplars and more abstract information. It seems intuitively attractive that the manner in which concepts are represented differs across domain. For example, it is easy to think of exemplars of many object categories. It is even possible to think of exemplars of such social categories as good or honest. However, Lingle, Altom, and Medin (1984) have pointed out that these exemplars seem smaller in number, less accessible, and less central to the concepts than are exemplars of object categories.

This distinction leads to the proposal of two distinct categorisation strategies. With exemplar based object categorisation, classification will be done on the basis of analogy to a concrete instance (Brooks, 1978; Estes, 1986; Nosofsky, 1984; Wattenmaker, 1993). As Wattenmaker (1995) points out, this also suggests the possibility that the relationship which holds between the features of an object to be categorised will also influence decisions. It is well known that the relationship which holds between features tends to be specified in, at least some, concepts (Goldstone, Medin, and Gentner, 1991; Medin, Goldstone, and Gentner, 1990). There is also evidence (Gentner, 1983; Goldstone, Medin, and Gentner, 1991) that relational properties influence both similarity calculations and the use of analogies. However, relational properties tend to inhibit the learning of linearly separable categories.

Accordingly, it is likely that a different strategy is used in social categorisation. This strategy is likely to be determined by the abstractness of social concepts. In general, features of social concepts tend to be more abstract than those of object concepts. So, in terms of the earlier example involving the academic who resigns from her job, we might classify this behaviour as altruistic. Such a behaviour is not part of our concept of altruism but we infer from the abstract features that are associated with the concept of altruism in order to make the classification. It has been suggested (Murphy and Medin, 1985; Rips, 1984) that much categorisation is based on inference rather than feature matching. That social categorisation, in particular, is inferential, explains why the causal reasoning
mechanisms used to resolve inconsistencies which were discussed earlier, play such a role in social classification decisions.

Based on this analysis Wattenmaker ran an extensive series of studies to test for interactions between domain specific knowledge structures and abstract category structures. He found that across a wide range of tasks and materials, linearly separable categories were more compatible with social than with object categories. In sorting tasks, subjects were more likely to sum characteristic features and form LS categories with social materials. In learning tasks, LS categories were easier to learn with social materials but NLS categories were easier to learn with object materials. He also found evidence that the integration strategies used in categorisation varied across domain. In object conditions subjects tended to focus on single dimensions, use configural properties and rely on analogy. In social conditions summing features and learning LS categories was the strategy used.

What are the implications of this work for cognitive algebraic and procedural models of the belief revision process? Firstly, it would seem that any model of the process has to be domain specific. Certainly cognitive algebraic models, with their reliance on averaging and summation, are. Also of importance is the fact that the vast majority of experimental studies examining belief revision in general, and order effects specifically, have used social stimulus materials. Not only are cognitive algebraic models of belief revision domain specific by virtue of Wattenmaker's argument, they are domain specific by virtue of the data which they are based upon.

Secondly, and more importantly, Wattenmaker's work casts serious doubt on the independence assumption. If the summation of features in social categorisation is an inferential process, then it is impossible to work out the subjective value of a piece of evidence independently of other evidence. It is probably the case that additive models seemed to account for so much of the data produced by workers in the area of impression formation not because the meaning of the adjectives which subjects received was derived independently, but because inconsistencies between the adjectives were resolved
inferentially. This kind of inference is best viewed as a form of pragmatic inference where the task is to constrain the possible meanings of a piece of information rather than to derive the only meaning which that piece of information may convey. If, as Asch and Zukier (1984) have suggested, causal chains of reasoning are implicated, then it is clear that early occurring adjectives provide a context for this pragmatic inference. Indeed, it could very well be argued that, in the case of recency effects due to the equalisation of subjects' attention, later occurring adjectives provide the context for inferences concerning the earlier occurring adjectives.

3.3.3.2 Background knowledge about correlated features

The second strand of research which is strongly suggestive of a problem for both cognitive algebraic and procedural models of belief revision is very strongly related to the research discussed in the previous sub-section. This is the work examining subjects' sensitivity to the correlations which exist amongst features of objects in the world. For example, Medin, Altom, Edelson and Fresko (1982) demonstrated that subjects were sensitive to feature correlations and that they used them in their categorisation judgements in experiments with novel categories. Malt and Smith (1984) found essentially the same result with natural categories. Such findings are predicted by Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976) who proposed that natural categories divide the world up according to clusters of features. They further claimed that basic categories maximise the correlational structure of the environment by preserving these feature clusters.

Murphy and Medin (1985) have claimed that Rosch et al's view of conceptual knowledge is inadequate. They point out that such an account fails to specify a mechanism by which the encoding of correlations is constrained. Neither does it explain why people perceive and store "illusory correlations" (Chapman and Chapman, 1967; 1969). They claim that the correlational account must be supplemented with a theory-based approach. Only in this way can the two major problems for the correlational account be solved.
Despite this qualification, there appears to be no doubt that subjects are sensitive to the rate at which some features of some phenomena co-occur. The discussion in the previous sub-section suggests that knowledge about correlation is most important in non-linearly separable categories. As discussed above, Wattenmaker has provided evidence that, with objects, categorisation judgements are made based on single or configural properties of the object to be categorised.

How does this finding sit with cognitive algebraic and procedural models of belief revision? Not very well in fact. If the previous sub-section was an attempt to show that cognitive algebraic models in particular were right about belief revision within social contexts, but for the wrong reasons, this sub-section will attempt to demonstrate that procedural models of belief revision, which offer the best hope for explaining belief revision with object materials, will ultimately be found wanting. As with cognitive algebraic models, the fatal flaw in procedural models is connected with the independence assumption. This will be demonstrated for the model of Hogarth and Einhorn, which is the best specified of existing procedural models.

That Hogarth and Einhorn's model suffers from the independence assumption is best demonstrated by reference to their contrast assumption. It will be remembered that the contrast assumption leads to predictions of large increases or decreases in belief when the absolute difference between the subjective scale value of new evidence is much less or much greater than background levels of belief. Accordingly, in many cases, the size of the recency effect is predicted to depend on this difference in subjective value. Tubbs, Gaeth, Levin and van Osdol (1993) have provided evidence that this is the case. Interestingly, they also found that difference in absolute strength was a greater determinant of their order effect than was the inconsistency or consistency of the evidence. To a large extent, the contrast assumption, and Tubbs et al's finding, makes sense. It very nicely captures the potential of very strong, or very weak, evidence for surprise. However, it is inadequate.

Hogarth and Einhorn's contrast assumption is based on differences in the relative strengths of new evidence and background beliefs. The surprise which they model is that
which comes from having a seemingly very strong (or very weak) hypothesis called into question by a very weak (or very strong) piece of evidence. It fails to take account of subjects' background knowledge about the world. Whether this background knowledge is in the form of theory and correlated features (Murphy and Medin, 1985) or correlated features alone, is of little interest here. What is of interest is the observation that, because humans store much information about objects in the world in terms of correlated features, that version of the independence assumption which is embodied in Hogarth and Einhorn's contrast assumption, is unworkable. As well as the surprise which comes from new evidence which is much stronger or weaker than the strength of currently existing beliefs, information about certain features of objects is also likely to lead to expectations concerning other features. If such expectations exist, then the independence assumption must be invalid. It will be remembered that the tenuous explanation given for the manner in which subjects used the information which they received in Experiment 1 also implicated expectation. The experiments contained in the next chapter will attempt to test both that explanation and the sufficiency of Hogarth and Einhorn's contrast assumption, in the light of what has been discussed in this sub-section.

Before moving on to discuss the final a priori reason for doubting the sufficiency of both cognitive algebraic and procedural models of belief revision, two further points should be made about correlated features. Firstly, as mentioned in the section on domain specific differences in the organisation of knowledge, correlated features are themselves a feature of object knowledge. Therefore, it is to be expected that the preceding discussion of correlated features, and their implications for models of belief revision, relates primarily to object knowledge. Secondly, also in the previous section, social categorisation was characterised as more inferential in nature than was object categorisation. It would be a mistake however to think that inference is not involved in object categorisation. For example, if correlated features are implicated in the process, then any expectations which information about one of the features leads to, must be derived inferentially. Likewise, the extent to which later information meets, exceeds, or fails to live up to, those expectations, must also be derived inferentially.
3.3.3.3 The importance of pragmatics

This section of the chapter will deal with two issues where pragmatics are likely to be very important. The first of these is the general issue of the pragmatic determinants of meaning, which will be very briefly discussed. The second issue is one which has received recent attention in the decision making literature. That is the question of how the importance of evidence is pragmatically determined.

3.3.3.3.1 Pragmatics and meaning

Just as the two preceding arguments against the sufficiency of current models of belief revision are related, so too does this argument relate to the other two. This argument comes from work done in the last ten years which questions the representational-computational approach to cognitive science. The questions which this work raises for traditional approaches to cognition are far reaching (Shanon, 1988; Edelman, 1992) and therefore, a thorough treatment of them is beyond the scope of this thesis. Instead, discussion will be confined to just one question - the relationship between pragmatics and meaning - and to a critique by just one worker - Benny Shanon (Shanon, 1988). The advantage of imposing such a constraint on the discussion is twofold. Firstly, discussion will be limited to that specific, and relevant, part of the more general argument. Secondly, although the conclusions which Shanon draws are by no means representative, the criticisms upon which he bases those conclusions are.

Shanon's critique is based on his argument that representations, as defined in the cognitive literature, cannot account for the knowledge manifested by the cognitive agent's behaviour (Shanon, 1988, pg. 71). He focuses on the semantic representation of the meaning of linguistic expressions and argues: (1) that semantic representations, even if defined in terms of a well specified code, cannot exhaust the meaning of linguistic and other expressions; (2) that even if a semantic representation could exhaust meaning, the fact that meaning is context dependent introduces the possibility of infinite numbers of meanings for linguistic expressions; (3) that contextual problems cannot be solved using a
two stage model where the first stage involves the representation of the basic core of the meaning of a linguistic expression.

Obviously, Shanon's argument has ramifications for all of cognitive science. To remedy the problems he has identified, he makes very specific suggestions concerning new foundations for the discipline. These suggestions, although contentious, are not at issue here. What is of most interest is how Shanon (1988; 1993) has applied his critique to existing work in cognitive science. Specifically, he analyses Tversky's (Tversky, 1977; Tversky and Gati, 1978) contrast model of judgements of similarity. Once again, space and relevance constraints forbid a thorough exposition of either Tversky's or Shanon's position here. Suffice it to say that Tversky's model has been very influential - Eysenck and Keane (1995) describe it as one of the most long standing models in cognitive psychology. Essentially the model claims that the similarity of two objects is based on some function of the attributes which the objects share minus the attributes exclusive to each. The basic approach is captured in Equation 3.11:

\[ s(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad \text{Eq. 3.11} \]

where \( a \) and \( b \) are two objects, \( s \) is the similarity of these objects, \( A \) is the set of attributes of object \( a \), \( B \) is the set of attributes of object \( b \), \( A \cap B \) refers to the set of attributes common to both objects, \( A - B \) refers to the features unique to \( a \), \( B - A \) refers to the features unique to \( b \), the function \( f \) weights some features in terms of their salience, and the parameters \( \theta \), \( \alpha \), and \( \beta \) reflect the relative importance of the common and distinctive feature sets. Because of the weighting function, Tversky's model does, like existing models, take into consideration the judgmental context. Some features of objects will be more important than others in certain contexts. However, as Shanon (1988) points out, the contrast model assumes that the definition of the features themselves is prior to, and independent of, the evaluation of similarity. This assumption is similar to the independence assumption contained in existing models of belief revision.
Shanon claims that, not only is the differential weighting given to certain features due to context, in some cases features are dependent on the evaluation of similarity itself. He gives the example (Shanon, 1988, pg. 75) of a new-born baby being inspected by aunts from different sides of the family. Aunts from each side will claim that the child is similar to a family member from her side of the family, and then proceed to find the evidence in the child's features to back up her claim. Thus, rather than the features being evaluated and similarity being computed based on the prior evaluation, similarity is first postulated and features are then chosen.

Much the same type of argument may be applied to models of belief revision. For example, in the previous sub-section, the importance of background knowledge about the relationship which holds between evidential features was pointed out. Current models of belief revision are inadequate in two ways. Firstly, they are based on traditional views of representation. This view is best expressed by Fodor (1975) who claims the existence of a language of thought which allows for a characterisation of all knowledge. Only by making this assumption can the independence assumption be justified. The continued holding of the independence assumption leads directly to the second inadequacy of existing models of belief revision. That is, they all lack a pragmatic component. Without such a pragmatic component, the meaning of a piece of evidence cannot be constrained. If social knowledge is best characterised by its flexibility, as the work discussed in the section before last suggests, then, for models of the belief revision in social settings, the problem becomes one of explaining how the meaning of a piece of evidence is constrained and only then, of explaining how that evidence comes to have an impact on the beliefs of the cognitive agent. As Sperber and Wilson (1986; 1995) have persuasively argued, constraining the meaning of a piece of information is an inferential process. Accordingly, the process of revising one's beliefs must also be seen as a process which involves pragmatic inference. The importance of pragmatics to any model of the belief revision process will be further illustrated in the next section.
3.3.3.3.2 Pragmatic determinants of evidential importance

The most essential feature which any alternative model of the belief revision process would need to possess is a pragmatic component. In addition to the argument made in the previous section concerning pragmatics and meaning, such a pragmatic component is necessary for a further two reasons. Firstly, although recent models of the process of belief revision have emphasised the importance of the anchor, only Lopes (1987) has speculated on how a piece of evidence comes to be used as an anchor. She suggests that the information which is most diagnostic, or important, for the organism, will anchor the judgement. However, the question of what makes information important has never been addressed in the belief updating literature. As was demonstrated in the first chapter of this thesis, the question of what makes information relevant is a central topic in contemporary work on reasoning. There are three accounts of information selection which include a well worked out pragmatic component (Oaksford and Chater, 1994; Sperber, Cara and Girotto, 1995; Evans and Over, 1996).

In addition to this recent work in the reasoning literature, there also exists work in the decision making literature which suggests that there are pragmatic determinants of the evidence which is considered to be important when attempting to decide between alternatives. For example, Shafir (1993) asked subjects to imagine that they were serving on a jury in an only-child sole-custody case. Subjects were also given a list of attributes possessed by each parent. Parent A was average in all respects, whereas parent B possessed both very positive, and very negative, attributes. When asked "To which parent would you award sole custody of the child?", 64% of subjects said they would award custody to parent B. However, when asked "To which parent would you deny sole custody of the child?", 55% of subjects said they would deny sole custody to parent B. Shafir explains these results in terms of reason-based choice. When asked to choose between alternatives subjects focus on those attributes of the options which are reasons for choosing them. Thus, more subjects choose parent B because parent B possessed more extremely positive attributes than did parent A. On the other hand, when asked to reject one of a pair of alternatives, Shafir claims that subjects focus on reasons for rejection.
Accordingly, as parent B also possessed more extremely negative attributes than did parent A, parent B was denied custody more often.

Similarly, Maachi (1994) has demonstrated that the base-rate fallacy (Tversky and Kahneman, 1980) where subjects fail to use base rate information when estimating the probability that an event will occur, is dependent on the wording of both the problem contexts and questions used. For example, she gave two groups of subjects the suicide problem (Bar-Hillel, 1980; Tversky and Kahneman, 1980). Subjects in the control group received the original version of the problem:

*Consider the following assumptions regarding suicide. In a population of young adults, 80% of the individuals are married and 20% are single. The percentage of deaths by suicide is three times higher among single individuals than among married individuals.*

*What is the probability that an individual, selected at random from those that had committed suicide, was single.*

Subjects in the experimental condition received exactly the same problem but the final question was changed to read:

*What is the probability that a suicide, selected at random from the population of young adults, was single?*

For the control condition, Maachi's results were very similar to those of Tversky and Kahneman. Sixty six percent of subjects in the control group failed to use the base rate. In her experimental condition, on the other hand, only thirty four percent of subjects fail to use the base rate.

In the suicide problem, the percentages refer to the base rates for single and married people in the general population. The ratio of deaths by suicide amongst single
versus married people is known as the likelihood ratio, and in this case is 3:1. Tversky and Kahneman explained the tendency amongst their subjects to use only the likelihood ratio in terms of a perceived causal link between being single and committing suicide.

Conversely, Maachi interprets their results as being due to the question which subjects were asked about the information. She claims that, as the question referred to the population of suicide victims, 75% of whom are known to be single, subjects may infer that the information concerning suicide victims has been produced by a consideration of base rate information. She claims as support for this interpretation the fact that the majority of subjects in her experimental condition take the base rate into account when producing their estimations. Crucial to this effect is the re-wording of the question where it is made clear that the likelihood information is independent of the base rates.

Both the Shafir and Maachi studies illustrate that pragmatic factors are crucial in determining the information that is important to the subject in any given context. They also suggest that information which is considered to be irrelevant in any given context may be ignored completely. Thus, not only would a theory of belief revision which possessed a pragmatic component, be able to predict which information was likely to be important to the organism in any given context, it would also have a mechanism for constraining the information which is used in making the judgement.

3.4. SUMMARY AND CONCLUSIONS

This final section of the chapter will attempt to do two things. Firstly, the preceding discussion will be summarised, and secondly, the conclusions which may be drawn from what has already been said will be discussed.

3.4.1 Summary

The first concern of this chapter was to review in some detail the literature on order effects in belief revision. It was argued that although the majority of such effects are of primacy, there are also a substantial number of recency effects in the literature. However,
recency may normally be attributed to some procedural variation rather than to some characteristic of the information which subjects are using to revise their beliefs. It was also argued that despite the extensive efforts to rule out explanations for primacy based on Asch's "change of meaning" hypothesis, Asch's account has not been definitely ruled out. This fact is due to the under specification of Asch's hypothesis. Specifically, it is not clear to what extent the process of belief revision works backwards, as well as forwards.

The algebraic and procedural models which have been put forward to both describe and explain the belief revision process were also reviewed. The biggest difference between the two types of models is that algebraic models attempt to describe the output of the judgement process whereas procedural models attempt to explain what people do when they integrate information in order to revise their beliefs. A second advantage of certain procedural models (Hogarth and Einhorn, 1992) is that they provide a rationale for the differential use of averaging and summation predicted by various algebraic models. Two further improvements made by procedural models in our understanding of the process were also discussed. The first of these is the assumption that subjects adopt an anchor when forming and revising their beliefs. The second is Hogarth and Einhorn's contrast assumption which states that the size of an order effect is dependent on the difference in subjective value between the pieces of evidence which the subject receives.

It was argued however, that despite these improvements, procedural, as well as algebraic, models are unlikely to provide an adequate characterisation of the belief revision process. This is mainly due to a failure to consider the importance of both pragmatics and structure in background knowledge to the process. Although both types of models agree that the context of judgement is important, they fail, because of the independence assumption, to appreciate the importance of evidential context. The independence assumption states that the subjective value of a piece of evidence is determined independently of any other evidence which the subject might receive and is also found in the "language of thought" argument, best expressed by Fodor (1975). Drawing on various arguments and findings from the literatures on concepts, pragmatics, and decision making it was argued firstly, that the independence assumption is
unwarranted and secondly, that a pragmatic component is necessary to any complete model of the belief revision process.

It was further argued that because of the failure of existing procedural models to take findings about the underlying structure of knowledge into account, advances such as the contrast assumption will ultimately prove fruitless. Likewise, it was argued that the notion of an anchor is not adequately specified. Once again, a failure to consider pragmatics was implicated here. None of the existing procedural models are capable of predicting what information will be used as an anchor. In addition to this, it was claimed that the inferential aspect of belief revision is ignored by existing models. Although it was argued that any account of the belief revision process is likely to be domain specific, inference is crucial to an understanding of the process in any domain.

3.4.2 Conclusions

The first conclusion which may be drawn from the preceding discussion is that there is no adequate model of the belief revision process currently in existence. That is not to say however, that existing approaches and models have no merit. For example, it would seem that both Asch and Anderson were correct in their intuitions. Based on the work of investigators such as Wattenmaker (1995) and Zukier and Asch (1984), social knowledge seems to be more abstract and inferential in nature, than is object knowledge. It would seem that on a task such as impression formation, the meaning of later adjectives is determined by earlier adjectives in terms of the inferences made about the later adjectives based on both the earlier adjectives and background knowledge. However, algebraic-type operations also characterise categorisation with social knowledge. The work which exists on linear separability makes many of the same predictions as investigators such as Anderson and Shanteau. Likewise, the more recent work on anchors and the importance of evidential contrast also seems to contain more than a germ of truth. However, because of this work's failure to incorporate ideas from the literatures on concepts and pragmatics, it must be seen as flawed.
There also remain several questions which previous work has not addressed. An extremely important, but as yet unaddressed, question, concerns the flexibility of the process. To what extent is the evaluation and integration of evidence reversible? Massaro and Friedman (1990) have characterised all existing models of information integration as being forward working. That is, once a piece of evidence is evaluated and integrated, it cannot be re-evaluated in the light of subsequent information. However, if it is accepted that primacy is the result of inference from background knowledge, then surely the recency effects which exist in the literature must be seen as a re-evaluation of earlier information in terms of the later information which subjects receive.

This point leads to the question of how best to characterise the role of forgetting in the process. If it can be shown that the evaluation of evidence is reversible, then it becomes likely that subjects are holding earlier information in working memory and using this information as the basis for mental operations (i.e. inferences) in conjunction with later occurring information and information from background knowledge. The notion of a mental model (Johnson-Laird, 1983; Gentner and Stevens, 1983) seems ripe for exploitation as an explanatory concept here. It is likely that the structure of such mental models will differ across domains, but undoubtedly some structure which allows for the operation of basic inferential processes on information from both the world and background knowledge is required.

Another inescapable conclusion from what has already been said is that the notion of focus has some part to play in any understanding of the belief revision process. Although focus is currently badly understood (Kroger, Cheng, and Holyoak, 1993; Love and Kessler, 1995), it plays an important role not just in theories of reasoning, but also in recent work on language understanding (e.g. Moxey and Sanford, 1993). Focus is likely to be important in understanding belief revision in two ways. Firstly, pragmatically determined focus is likely to underlie the selection of an anchor in any particular context. Secondly, focus will determine the particular hypothesis or alternative to which subjects attend. In short, it seems highly probable that the notions of pragmatically determined
focus and mental models have the potential to be very powerful tools in aiding understanding of the belief revision process.

However, before the theoretical constructs, discussed above, may be applied to the phenomenon of belief revision, more experimental work is required. Accordingly, the following two chapters will have the joint aim of exploring the effect found in Experiment 1, and of shedding light on the process of belief revision. The next chapter will start with a re-consideration of that effect in the light of what has been said in this chapter. It will relate the notion of expectation (postulated in Chapter 2 to explain the effect of evidence on subjects' beliefs about the alternatives present in the experimental scenarios) to Hogarth and Einhorn's contrast assumption. Finally, it will attempt to test the adequacy of that assumption in the light of issues discussed in this chapter.
4.1. INTRODUCTION

This chapter will consist of the details of three experiments designed to further examine the results of Experiment 1. There were two aspects to Experiment 1. The first of these was the information which subjects selected to enable them to decide between the alternatives. Experiment 2 was designed to further examine the factors involved in information selection. The second aspect of Experiment 1 was the use which subjects made of the information which they received. This chapter constitutes an attempt to further explore how subjects use information to decide between alternatives. A second aim of this chapter will be to shed further light on the processes involved in belief revision in general. The area of belief updating was critically reviewed in the previous chapter. Accordingly, the result of interest from Experiment 1 will be reconsidered in light of the previous discussion.

4.1.1 Another look at Experiment 1

It will be remembered that Experiment 1 employed Doherty et al's (1979) pseudodiagnosticity task in a novel way. As well as asking subjects to choose information to decide between alternative hypotheses, subjects were given the information which they had selected, and asked to use that information to state which of the hypotheses was most likely. All subjects received a single scenario such as the following:

Your sister has just bought a new car. It's either a model X or a model Y but you can't remember which. You do remember that it has four doors and a top speed of over 95 mph.
The scenario above may be broken down into two components. The first of these are the evidential features (having four doors and a top speed of over 95 mph) and the second are the alternatives amongst which the subject must decide.

As one of the aims of Experiment 1 was to examine the effect of rarity on subjects' choice and use of information in deciding between alternatives, there were two sets of features for each scenario. In the example given above, both of the features have a high subjective probability. That is, the majority of cars have four doors and a top speed of over 95 mph. The second feature set consisted of one feature with high subjective probability and one of low subjective probability. So, in the example given above, such a set might be having four doors and a top speed of over 165 mph.

Following the scenario, subjects were told the percentage of one of the models of car which possessed one of the features (in all cases relevant to this discussion they were told that this percentage was either sixty five or ninety five). Next, subjects were asked to choose one piece of information from amongst the remaining three to help them to decide whether their sister's car was most likely to be a model X or a model Y. For the example given, the four possible pieces of information are given below:

A. The percentage of model Xs which have four doors

B. The percentage of model Ys which have four doors

C. The percentage of model Xs which have a top speed of over 95 mph

D. The percentage of model Ys which have a top speed of over 95 mph

In all cases the first piece of information which subjects received was A, so their choice was restricted to one of B, C, or D. For the feature sets consisting of one rare and one common feature, A always concerned the rare feature. For the feature sets consisting of two common features, A always concerned the feature not shared by both sets. For
example, in the scenario given above, for the common feature set A always concerned the percentage of model X cars said to have a top speed of over 95 mph.

The evidential impact of the piece of information which subjects chose was manipulated so that it was either positive or negative relative to the X hypothesis. To achieve this manipulation the percentages of B, C, and D were set at either 25%, 75%, and 25% (positive impact on hypothesis X) or 75%, 25%, 75% (negative impact on hypothesis X). In this way, regardless of the piece of information which any subject chose, the sign of that piece of evidence's impact could be controlled for.

The results of this part of Experiment 1 are depicted in Figure 4.1. It will be remembered from the discussion of those results that there was a significant interaction in subjects' confidence ratings between feature set (two common features or one rare and one common), evidence selection (diagnostic or pseudo-diagnostic), and feedback (either positive or negative relative to the focal hypothesis). Subsequent analysis revealed that the overall significance of this result could be attributed to a significant difference between two of the means involved in the interaction. Both of the means involved were produced by subjects who chose pseudo-diagnostic information and received positive feedback from
their selection. The only difference between the groups was that one of them initially received information about a common feature whilst the other initially received information about a rare feature. Those subjects in the latter condition displayed more confidence in the Y(or non-focal) hypothesis (mean of 45 on a scale where 0 corresponded to complete certainty that Y was the case and 100 corresponded to complete certainty that X was the case) whilst those subjects who initially received information about a common feature were highly confident that X was the case (mean confidence of 83).

This result was surprising as, on the face of it, those subjects who initially received information about a rare feature might be thought to have possessed stronger evidence in favour of hypothesis X than did subjects who received information about a common feature. This is because a rare feature of an object is more diagnostic than is a common feature. In the car scenario above, knowing that the target car possesses a very rare feature in common with the majority of instances of its hypothesised category would seem to be very strong evidence that the hypothesised category is the correct category.

One possible explanation for the counter-intuitive results of Experiment 1 is that expectation played a role in subjects' expression of their confidence. As was discussed in Chapter 2 such an explanation would claim that the first piece of evidence which subjects received led to the formation of expectations about the second piece of evidence. When those expectations were not met, subjects lost confidence in the focal hypothesis. Of course, this account is based on the assumption that subjects are revising their beliefs about the hypotheses in the light of new information. In Experiment 1 subjects received the information in a sequential fashion which should lead to the revision of their initial beliefs in the light of later information. Unfortunately however, subjects were only asked to express their relative confidence in the alternatives upon receipt of both pieces of evidence. Accordingly, based on the results of Experiment 1, it cannot be claimed with certainty that subjects are adjusting their beliefs in the light of the second piece of evidence. One of the purposes of the first study to be described in this chapter will be to establish that subjects are revising their beliefs in the light of new information.
4.1.2 Examining the role of expectation using the order effects paradigm

It was suggested in the discussion of Experiment 1 that a useful test of the role which expectation played in the effect found in that study could be carried out using the order effects paradigm. The effect in Experiment 1 was found when a strong piece of evidence (concerning a rare feature) was followed by a weaker piece of evidence (concerning a common feature). If subjects' expectations underlie their interpretation of the second weaker piece of evidence, then when the same information is presented in the reverse order, the result of Experiment 1 should be reversed. This is because subjects receiving the weaker piece of evidence first will have lower expectations about the second piece of evidence. Accordingly, their confidence in the focal hypothesis should increase upon receipt of the extra information, whereas in the strong/weak order their confidence will decrease. The advantage of using the order effects paradigm is that it allows for the examination of subjects' relative confidence in the hypotheses upon receipt of each of the pieces of information.

In terms of the literature on order effects discussed in Chapter 3, the order effect predicted here is one of recency. In other words, subjects who receive the stronger piece of evidence second will have greater confidence in the focal hypothesis than will subjects who initially receive the stronger piece of evidence. It is important however to distinguish between the reasons for this prediction and those of Hogarth and Einhorn's contrast assumption which would also predict recency. In this case the contrast assumption predicts recency because of the difference between the subjective values of the two pieces of evidence. Weak evidence followed by strong evidence always results in upwards revision of belief, whereas strong evidence followed by weak evidence always results in downwards revision of belief. The claim made here however, is that the subjective value of the later occurring evidence will, in certain cases, be determined by the subject's expectations about that piece of evidence. These expectations are derived from both the early occurring evidence and background knowledge.
Although the aim of Experiment 3 is simply to demonstrate the existence of an order effect with materials similar to those used in Experiment 1, Experiments 4 and 5 will attempt to demonstrate that, in certain cases, Hogarth and Einhorn's contrast assumption will be an inadequate predictor of when order effects will, and will not, occur. Instead, it will be argued that the role of expectations in the interpretation of evidence must be acknowledged.

4.2. EXPERIMENT 3

4.2.1 Introduction

Experiment 3 will be a direct test of the expectation-based account given for the results of Experiment 1. To test this account an order effects paradigm will be used where subjects receive the same information but in different orders. In addition to the order effects manipulation there will be several changes to the procedure used in Experiment 1. Firstly, subjects will be asked to express their relative confidence in the hypotheses upon receipt of each piece of information. Such a procedure allows for the examination of both the first and second pieces of evidence on subjects' beliefs about the hypotheses. Secondly, because there was no independent validation of the perceived frequency of the evidential features used in Experiment 1, validation will be obtained in this experiment by asking subjects about their expectations concerning the second piece of evidence once they have received the initial piece of information. This will enable a manipulation check on the perceived frequency of the evidential features used.

There are several predictions which may be made about the results of this experiment in the light of what has previously been said about the role of expectations in the belief revision process. Firstly, the effect of manipulating evidence order will be to produce a recency effect. This is because the order manipulation will determine which evidence subjects receive first. Those subjects who receive the stronger piece of evidence first will have high expectations about the second weaker piece of evidence. Because this second piece of evidence will not meet those expectations, its effect will be to leave
subjects' beliefs unchanged, or, to cause them to revise downwards their confidence. On the other hand, those subjects who receive the stronger piece of evidence second will not have high expectations. Accordingly, the second piece of evidence will cause them to revise their belief upwards. This leads to the second prediction about the results of Experiment 3. Subjects who receive the weak evidence first will revise their beliefs upwards significantly more often than will subjects who receive the strong evidence first. Finally, it is predicted that subjects who receive the strong evidence first will have significantly higher expectations concerning the second piece of evidence than will subjects who receive the weak evidence first.

4.2.2 Method

Subjects: the subjects were 60 students attending an Open University undergraduate summer school in cognitive psychology. 44 of the subjects were female and 16 of them were male. Their mean age was 40.3 years. The oldest subject was 68 and the youngest was 23.

Materials: each subject received a handout that comprised an instruction sheet and three problems. The instructions were as follows:

Accompanying these instructions is a series of three decision problems which require you, in the light of certain information, to rate your confidence in two alternatives and to answer some simple questions. Detailed instructions on what to do are contained in the problems but it would be very helpful if you could bear the following general points in mind throughout the experiment.

(1) Read each problem carefully and think hard before responding.

(2) Although there are some surface similarities between the problems, they all concern different scenarios, so you should think carefully about each scenario.
(3) Don't refer back to previous problems when working on the later problems.

(4) Each problem consists of two pages which are stapled together. It is important that you receive the information they contain in the intended order. Don't go on to a later page until you have finished the one you are currently working on.

If you have any questions at any time, please raise your hand and the experimenter will answer them.

Once again, the structure of the problems used was taken from Mynatt, Doherty and Dragan (1993). The first page of each problem given to subjects in the first condition was as follows:

Your friend is an engineer and works for a large construction company. It's either company A or company B, but you can't remember which. You do remember that he drives a company car and that he earns over £50,000 a year.

You have the following piece of information:

95% of engineers working for company B earn over £50,000 a year.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend works for company A and the other to complete certainty that your friend works for company B. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.
certain that your friend works for company A
certain that your friend works for company B

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of engineers working for company A would you expect to earn over £50,000 a year?

(2) What percentage of engineers working for company B would you expect to drive a company car?

(3) What percentage of engineers working for company A would you expect to drive a company car?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

As may be seen from the above problem materials, subjects were asked about their expectations for each possible piece of evidence. This was done so as not to alert them to the piece of evidence which they would receive. Only their answers about the piece of evidence which they were to receive will be analysed here. The second page of each problem contained the following information:

Here is a second piece of information about the problem on the previous page:
70% of engineers working for company B drive a company car.

Once again, there followed a rating scale where subjects were asked to indicate their confidence in each of the hypotheses.

Subjects in the second condition received exactly the same problems except that the order in which they received the probabilistic information was reversed. Thus, those subjects in the second condition received the information that 70% of engineers working for company X drove a company car on the first page of the problem and on the second page the information that 95% of engineers working for company X earned more than £50,000 a year.

There were three different contexts used. The second concerned a friend staying in one of two hotels in Paris whose room costs £165 a night and has an en-suite bathroom. The third context involved a friend who lives on one of two streets, in a house with a swimming pool and a garage. Examples of materials using each of these contexts are given in Appendix 1. The two features about which information was given (company car and salary in the engineer problem; garage and swimming pool in the house problem; bathroom and price in the hotel problem) were chosen so that one of them was a statistically frequent feature and the other was statistically infrequent. Accordingly, the infrequent features will be referred to as rare (salary of £50,000+, swimming pool, and rate of £165 a night), whilst the infrequent features will be referred to as common (company car, garage, and en-suite bathroom).

For ease of reference, the hypothesis about which both pieces of information were given will be referred to as the focal hypothesis and the second hypothesis, about which subjects were given no information, will be referred to as the alternative hypothesis. The confidence measure taken before subjects received the second piece of information will be referred to as confidence before, and the confidence measure taken afterwards, as confidence after.
Design: there were three predictions made in the introduction to this experiment. This first of these involved subjects' ratings of their relative confidence in the hypotheses (this will be referred to as Rating). This part of the experiment had a 2x2x3 mixed design. The between subject variable was Order (whether subjects received the strong piece of information first or second), whilst the within subjects variables were Before/After (subjects expressed their confidence in the hypotheses both before and after receipt of the second piece of information) and problem content (there were three different contents used - engineer, house and hotel).

The second prediction involved the direction in which subjects would revise their beliefs after receipt of the second piece of evidence. This will be referred to as Direction of Revision. This part of the study had a 2x3 mixed design. The between subjects variable was order, whilst the within subjects variable was problem content. The third part of the study involved subjects' expectations (after receipt of the first piece of evidence) about the second piece of evidence they were to receive. This dependent variable will be referred to as Expectations. This part of the study had a 2x3 mixed design. As before the between subjects factor was Order, whilst the within subjects factor was problem content. The design of the experiment is summarised in Table 4.1.

<table>
<thead>
<tr>
<th>Task</th>
<th>Independent Variables</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratings</td>
<td>1. ORDER</td>
<td>2 - STRONG 1ST;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>STRONG 2ND</td>
</tr>
<tr>
<td></td>
<td>2. BEFORE/AFTER</td>
<td>2 - Confidence Before</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confidence After</td>
</tr>
<tr>
<td></td>
<td>3. PROBLEM CONTENT</td>
<td>3. Engineer, House, Hotel</td>
</tr>
<tr>
<td>Direction of Revision</td>
<td>1 and 3</td>
<td></td>
</tr>
<tr>
<td>Expectations</td>
<td>1 and 3</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: The independent and dependent variables involved in Experiment 3.

Procedure: subjects were run in two groups of 25 and one group of 10, and were randomly assigned to one of the two order conditions. They worked on their problems individually. The order in which subjects attempted each problem was manipulated so that 5 subjects in each condition received the problems in each of their six possible orders.
4.2.3 Results

4.2.3.1 Confidence Ratings

Two subjects failed to provide confidence ratings on at least one of their six scales, so for the purposes of this analysis they will be excluded. This left 29 subjects in each of the between subjects conditions.

Subjects' confidence measures were analysed using a 2x2x3 mixed design analysis of variance. The between subjects factor was Order (this had two levels - Strong 1st and Strong 2nd), whilst the within subjects factors were Before/After (Confidence Before and Confidence After receipt of the second piece of information) and problem content. The means and standard deviations from this analysis are given in Table 4.2, whilst a full Anova table is given in Appendix 2.

It will be remembered that subjects were asked to express their confidence on a 100mm line. At one end of the line was labelled "certain...X" and at the other end "certain...Y". For the purposes of the analysis which follows subjects' markings on the line were converted to scores on a 100 point scale ranging from 1 to 100. Each point on this scale corresponds to one millimetre on the line and the higher is a subject's score on this scale the more confident she is that the X hypothesis is the case.

<table>
<thead>
<tr>
<th>Order</th>
<th>Before</th>
<th></th>
<th></th>
<th>After</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Engineer</td>
<td>House</td>
<td>Hotel</td>
<td>Engineer</td>
<td>House</td>
<td>Hotel</td>
</tr>
<tr>
<td>Strong 1st</td>
<td>63 17</td>
<td>61 15</td>
<td>64 21</td>
<td>58 17</td>
<td>64 14</td>
<td>59 18</td>
</tr>
<tr>
<td>Strong 2nd</td>
<td>57 15</td>
<td>60 16</td>
<td>63 14</td>
<td>71 17</td>
<td>73 17</td>
<td>72 18</td>
</tr>
</tbody>
</table>

Table 4.2: Means (in bold) and standard deviations from the analysis of subjects confidence ratings in Experiment 3.

The main effect for Before/After was significant $F(1, 56) = 9.72$, $p < .005$, as was the interaction between Before/After and Order $F(1, 56) = 21.01$, $p < .00005$. Mean
confidence before receipt of the second piece of evidence was 61, whilst the mean rating after receipt of that evidence was 66. There was no significant main effect for either Order

<table>
<thead>
<tr>
<th>Before/After</th>
<th>Strong information 1st</th>
<th>Strong information 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence before</td>
<td>63 (17)</td>
<td>60 (15)</td>
</tr>
<tr>
<td>Confidence after</td>
<td>61 (16)</td>
<td>72 (18)</td>
</tr>
</tbody>
</table>

Table 4.3: Means and standard deviations (in parentheses) involved in the significant interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 3.

F(1, 56) = 2.21, p = .143, or problem content F(2, 112) = .637, p = .531. Neither were any of the other interactions significant. The significant interaction between Order and Before/After is displayed in Figure 4.2, whilst the means involved are given in Table 4.3.

Tukey HSDs revealed significant differences between the means involved in the interaction. Subjects who received the weaker piece of information first were significantly more confident, on the second confidence measure, that the focal hypothesis was the case than they were on the first measure p<.001. They were also significantly more confident on this measure than were subjects in the strong information first condition on the first measure p<.001, and the second measure p<.001.
4.2.3.2 Direction of Revision

The number of subjects, broken down by Order and problem content, who revised their initial confidence rating upwards or downwards, as well as the number of those who made no revision, are given in Table 4.4. A quick look at the table confirms that, for all problem contents, there were more upwards revisions of belief when the strong evidence was received second than when it was received first. For all problem contents, when the strong evidence was received second, there were more upwards revisions of belief than downwards revisions. Similarly, for two out of the three problem contents, when the strong information was received first, there were more downwards revisions than upwards revisions. For each problem content, differences in the number of upwards revisions between conditions were analysed using Chi squares. For the engineer problem there were significantly more upwards revisions when the strong information was received second than when it was received first (χ² (1) = 4.5, p < .05). With both of the other problem contents the Chi squares produced insignificant results (χ² (1) < 3.84, p > .05 in both cases). So, although the pattern of upwards revision was in the expected direction, for two out of three of the problem contents, the difference between the conditions was not sufficiently large to produce a significant difference.

4.2.3.3 Expectations about the second piece of evidence

A total of 24 subjects failed to express their expectations about the second piece of evidence on at least one of their problems, so for the purposes of this analysis they were
excluded. Of the remaining 36 subjects, 20 received the strong piece of information first, and 16 received the strong piece of information second.

A 2x3 mixed design analysis of variance was used to analyse subjects' expectations about the second feature in relation to the focal (X) hypothesis. The between subjects factor was Order (although as subjects had received only one piece of information when they were asked this question, it is more precise to say that the between subjects factor was what information subjects had when asked the question). The within-subjects factor was problem content. A full Anova table for this analysis is given in Appendix 2.

There was a significant main effect for Order $F(1, 34) = 56.638, p < .0001$, but not for content $F(2, 68) = 2.054, p = .136$. There was also a significant interaction between Order and content $F(2, 68) = 6.717, p < .01$. Subjects' mean responses are presented in Table 4.5, whilst the interaction is displayed in Figure 4.3. Tukey HSDs for unbalanced

<table>
<thead>
<tr>
<th>Problem content</th>
<th>Strong information first</th>
<th>Weak information first</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineer</td>
<td>76 (25) N = 20</td>
<td>48 (29) N = 15</td>
</tr>
<tr>
<td>House</td>
<td>93 (15) N = 20</td>
<td>30 (25) N = 15</td>
</tr>
<tr>
<td>Hotel</td>
<td>90 (19) N = 20</td>
<td>51 (29) N = 15</td>
</tr>
</tbody>
</table>

Table 4.5: The means and standard deviations (in parentheses) involved in the significant interaction between Order and problem content from the analysis of subjects' expectations in Experiment 3.

Figure 4.3: The significant interaction between Order and problem content from the analysis of subjects' expectations in Experiment 3
samples revealed significant differences between means. Within problem contents the mean expectations for the second piece of evidence when the strong piece of evidence was received first differed significantly from mean expectations when the weak piece of evidence was received first (p < .001 for the house and hotel problems, p < .01 for the engineer problem). Within the Order conditions there were no significant differences between any of the problem content means although the difference between the means of the house and hotel problems when the weak piece of information was received first did approach significance (p = .058).

4.2.4 Discussion

The results of Experiment 3 are very clear. Two of the three predictions made at the outset received strong support, whilst the remaining prediction received partial support. As predicted, there was an overall recency effect in subjects' confidence ratings. Those subjects who received the strong information second were significantly more confident that the X (or focal) hypothesis was the case than were subjects who received the strong information first. Interestingly, only those subjects who received the strong information second revised their beliefs about the hypotheses to a significant degree. Whilst there was a significant difference between the initial and final confidence ratings of those subjects who received the strong piece of information second, there was no such significant difference between the ratings of those subjects who received the information in the reverse order.

Also interesting is the difference between subjects' confidence ratings after receipt of the first piece of information, and their expectations based on that information. As predicted, there was a highly significant difference between the expectations of those subjects who received the strong information first and those who received the strong information second. However, there was no such significant difference between the groups for initial confidence ratings. Although this latter result is somewhat surprising, it may be the case that subjects were waiting for more information before anchoring their judgement.
The possibility that subjects may use more than one piece of information as a basis for an anchor is discussed by both Lopes (1987) and Hogarth and Einhorn (1992). The fact that there was a significant difference between the final confidence ratings of the groups does suggest that subjects were using the scale to express their beliefs about the hypotheses.

The only prediction which did not receive full confirmation from the data was that concerning the effect of information order on the direction in which subjects would revise their beliefs upon receipt of the second piece of evidence. However, for one of the problem contents there was a significant difference in the direction predicted and for each of the other two problem contents the difference, although not significant, was also in the predicted direction. A closer examination of the data presented in Table 4.3 reveals that although a large majority of subjects who received the strong piece of information second did revise their confidence in the focal hypothesis upwards as was predicted, the same was not true of those subjects who received the strong information first. This result makes sense in terms of the interaction between the Order and Before/After manipulations. Subjects who received the stronger piece of evidence second were significantly more confident that the focal hypothesis was the case upon receipt of that second piece of evidence. However, subjects who received the stronger piece of evidence first were not significantly less confident after receipt of the second piece of evidence than they were before receipt of that evidence. Indeed, from the analysis of the direction in which subjects revised their beliefs, it seems likely that the weaker piece of evidence was not sufficiently weak, when received second, to cause a significant majority of subjects to revise their beliefs downwards.

Although this asymmetry in the results is initially surprising, it does not discriminate between the two opposing explanations for these results. It will be remembered from the discussion of Hogarth and Einhorn's contrast assumption that when strong evidence is followed by weaker evidence, the new belief is calculated as the old belief times the difference between the subjective value of the new evidence and the old belief. On the other hand, when strong evidence follows weak evidence the new belief is calculated as one minus the old belief times the difference between the value of the new
evidence minus the old belief. Accordingly, even though there was no significant
difference between the initial confidence ratings of either order condition, there is no
information available about the size of the difference between the subjective value of
either the weak or the strong evidence and initial belief strength. Therefore, there is no way
in which specific predictions about the magnitude of the revision may be made using the
contrast assumption.

The expectation account also sits comfortably with these results. The key in the
expectation explanation is firstly, whether expectations are met by the evidence and
secondly, the size of the difference between the expectation and the evidence. It is clear
that when the second piece of evidence was strong, it greatly exceeded subjects'
expectations. When the second piece of evidence was weak, as pointed out above, it is not
clear how great was the difference between expectations and the actual value of the
evidence. Therefore, although this experiment may be said to have provided some
evidence that expectations are important in the belief revision process, the experiment
itself was incapable of providing discriminatory evidence. The next experiment will
constitute an attempt to discriminate between the competing accounts.

4.3 EXPERIMENT 4

4.3.1 Introduction

The results of the previous experiment suggested the importance of expectations
from background knowledge to the process of belief revision. However, the results of
Experiment 3, although suggestive, may also be explained in terms of Hogarth and
Einhorn's contrast assumption. The purpose of this experiment is to demonstrate that the
contrast assumption is an inadequate explanation for order effects in belief revision. Before
going on to discuss exactly how this demonstration will be achieved, it will be useful to
say some more about how an explanation based on the contrast assumption differs from an
explanation implicating expectation.
The primary difference between the two explanations is in how well they capture evidential context. As stated previously, Hogarth and Einhorn's contrast assumption based model constitutes an advance on previous models of the belief revision process because it attempts to account for the context in which a piece of evidence is received. It does this by invoking differences in the subjective value of pieces of evidence. Accordingly, strong evidence followed by weak evidence always produces downwards revision. When the evidence is in the reverse order, upwards revision will always be observed (the role played by the sign of the evidence, i.e. which hypothesis the evidence supports, is moot. See overleaf for a discussion of this issue). In terms of the equations in which they express their model, the contrast assumption is captured by differential weighting of the subjective value of the new evidence. This differential weighting is dependent on the relative strengths of the subject's background belief and the new evidence. Thus, when the new evidence is weaker than background belief, the subjective value of the new evidence is weighted by the anchor. Conversely, when the new evidence is stronger than background belief, the weight is inversely proportional to the anchor.

This intuition appears to have been given support by a set of experiments reported by Tubbs, Gaeth, Levin and van Osdol (1993). These authors demonstrated both that there were recency effects with inconsistent evidence, and that the size of those recency effects was dependent on the degree to which the pieces of inconsistent evidence differed in subjective value. They also claimed that their finding of recency effects with evidence which had the same evidential sign (i.e. all positive or all negative as regards the focal hypothesis) but different subjective values (i.e. amongst positive evidential items some evidence was more positive than other evidence, and amongst negative items some evidence was more negative then other evidence) suggested that it was difference in subjective value rather than difference in sign which was the main factor underlying recency. Tubbs et al argued that this last finding was in direct contradiction to Hogarth and Einhorn's contrast assumption (which they claimed only predicts recency for inconsistent evidence).
There are two points which may be made about Tubbs et al's finding and their interpretation of that finding. Firstly, it is not at all clear that Tubbs et al's results are in contradiction with Hogarth and Einhorn's model. In that model, the source of differential weighting is the relationship between background belief and the subjective value of any new evidence. As discussed above, Hogarth and Einhorn define the case of new evidence with a subjective value greater than background belief, and the case of new evidence with a value smaller than background belief. Although in defining these two cases they place primary emphasis on the sign of new evidence, such an emphasis does not preclude consideration of differences in subjective value between two pieces of evidence.

The second point which may be made about Tubbs et al's results is that they are indicative of the problem with all existing models of belief revision, that is, the problem of what constitutes contradictory evidence. For example, when trying to decide which of two hypotheses is the case, it is not always clear whether evidence is marked for one hypothesis or the other. Imagine being asked to categorise a car which possesses four doors and leather upholstery as either a model X or a model Y. Does being told that 55% of model Xs possess four doors constitute evidence for, or against, the hypothesis that the target car is a model X? Both Lopes (1987) and Hogarth and Einhorn (1992) identify this problem of interpretation as a stage in the belief revision process where error can occur. However, neither of their approaches to the process can solve the problem. The situation is complicated when the subject receives two pieces of evidence. How does the subject decide whether these pieces of evidence contradict each other? Hogarth and Einhorn's unsatisfactory reply is to claim that if there is a difference between the subjective value of the new evidence and background belief (which itself must have been formed by earlier evidence), then that new evidence contradicts earlier evidence. The problem is, of course, one of working out the subjective value of a piece of evidence. It should be pointed out here that the term subjective value is used by almost all workers in the area of belief revision to refer to the value which a piece of evidence has for the subject. Although the term evidential value might be preferable, it will not be used here for two reasons. Firstly, and most obviously, the term subjective value will be used to preserve continuity with the previous literature. Secondly, the use of the term evidential value could lead to confusion.
as it could, in certain circumstances, be understood as referring to the epistemic utility (see Chapter 2 for a discussion of this concept) of a piece of evidence.

An account of the belief revision process which implicates expectation, presents a solution to many of the problems of working out the subjective values of pieces of evidence. In many cases where just one piece of evidence is presented, the problem of determining which hypothesis it supports is solved by inferring from background knowledge (in this instance about the rate at which models of cars possess various features) whether the hypothesised category possesses the evidential feature at a rate which is higher or greater than would be expected. Even in many cases where subjects receive sequential evidence, later evidence can have an expected value which is determined from both background knowledge and the earlier evidence. In the cases where these expected values exist, the actual value of the later evidence may be compared with the expected value and a decision about the sign of the evidence may be made.

This difference leads to a direct test of the two approaches. Consider the evidence which subjects received in Experiment 3. Asked to decide whether a target house which possessed a garage and a swimming pool was most likely to be on street X or street Y, subjects were told that 95% of houses on street X possessed a swimming pool and 70% possessed a garage. The expectation account of the results of Experiment 3 rest on the assumption that these features are in some way related. That is, given information about one of the features, subjects are able to derive expectations about the other feature from background knowledge. This assumption about features of objects being in some way related seems a reasonable one, especially in the light of the experimental work concerning correlated features and categorisation (e.g. Medin, Altom, Edelson and Fresko, 1982; Malt and Smith, 1984; Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976; Murphy and Medin, 1985) discussed in Chapter 3. Indeed, there is evidence from the literature on expertise in judgement (Phelps and Shanteau, 1978) that because of intercorrelations between features, experts are able to use more information than is usually thought.
However, the relationship which holds between the evidential features used in Experiment 3 is much simpler than one of correlation. Consider one of the pairs of features which were used in that experiment: having a swimming pool and having a garage. Although one would expect a house which possesses a swimming pool to possess a garage, one would not expect a house with a garage to possess a swimming pool. If these features were correlated, possession of any one feature would enable a prediction about the likelihood that an object possessed the second feature. In the case of the feature pair above, only one feature is predictive of the other.

Now consider a situation where subjects are given the same problem but this time receive information that 95% of houses on street X are over 200 years old and 70% have a garage. Neither of these features is likely to be predictive of the other. Accordingly, given one piece of information, subjects will not be able to derive expectations concerning the other piece of information from background knowledge. However, it is also likely that the probability that a house will possess a swimming pool is equal to the probability that a house is over 200 years old. Both features are rare in the general population of houses. Thus, they are equally diagnostic and will have the same independent subjective value for the subject. The anchor and adjustment model of Hogarth and Einhorn would, therefore, predict exactly the same results with the second set of evidential features as with the first.

The predictions of an account which implicates expectation in the process are very clear. Firstly, it is predicted that in the pre-test to establish both the perceived base rates of individual features, and the degree of perceived association between pairs of features, possession of the rare features used in Experiment 3 will be perceived to predict possession of the common features used in Experiment 3. Secondly, it is predicted that although the results of Experiment 3 will be replicated with pairs of associated features (i.e. feature pairs where one predicts the other, but not vice versa), there will be no order effect when subjects are given evidence concerning pairs of non-associated features. Simply put, an account which implicates expectation predicts recency with, for example, associated feature pairs of houses such as having a swimming pool and having a garage, but not with non-associated feature pairs such as having a swimming pool and being over 200 years...
old. On the other hand, Hogarth and Einhorn’s belief adjustment model predicts recency with both associated and non-associated features.

4.3.2 Pre-test

As the rationale for the pre-test has already been described this section will consist of a straightforward method and results section. Before describing the pre-test it should be noted that it was decided to abandon one of the scenarios used in the previous experiment (the hotel scenario) and to replace it with another scenario concerning a car belonging to the subject’s sister. This was decided because the hotel scenario was not considered to afford the possibilities for manipulation necessary for the planned experiment.

There were two parts to the pre-test the first of which was described in Chapter 2 and the second of which will be described in detail here. The first part was designed to establish the perceived frequency or base rate of certain features of objects. The second part sought to establish the perceived association between the most rare of those features and other features of the same objects. The features chosen from the first part of the pre-test will be presented before going on to describe the second part of the pre-test.

4.3.2.1 Establishing the base rates

For this experiment, and the two which follow, it was decided to use three problem

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Feature</th>
<th>Estimated Base Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOUSE</td>
<td>Possession of a swimming pool</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>Possession of a garage</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td>Being built between 1945 and 1985</td>
<td>69%</td>
</tr>
<tr>
<td>CAR</td>
<td>Having of a top speed of over 165 mph</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>Possession of a radio</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>Having four doors</td>
<td>58%</td>
</tr>
<tr>
<td>ENGINEER</td>
<td>Earning over £60,000 pa</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>Having a company car</td>
<td>62%</td>
</tr>
<tr>
<td></td>
<td>Working in general construction</td>
<td>52%</td>
</tr>
</tbody>
</table>

Table 4.6: The estimated base rates of the features selected for inclusion in Experiment 4
Two of these were used in Experiment 3 (the engineer scenario and the house scenario). The hotel scenario of the previous experiment was replaced (for the reason given above) with a problem concerning a car. Based on the results of the first part of the pre-test three features were selected for use in Experiment 4. It will be remembered from Chapter 2 that this part of the pre-test involved asking subjects to estimate how many objects, out of every ten thousand, would possess certain features. The features selected, and their estimated base rates are presented in Table 4.6.

4.3.2.2 Establishing the level of perceived associations amongst the features

Method

Subjects: 30 subjects participated in the pre-test. 7 of these were male and 23 were female. Subjects were drawn from the same population as that used for the first stage of the pre-test. Their mean age was 24.7. The youngest was 19 and the oldest was 48. Once again all subjects were paid for their participation in the experiment.

Materials: as with the first part of the pre-test, subjects were presented with a four page booklet consisting of an instruction sheet and three question sheets. The instructions were as follows:

On the following pages you will find a series of three questions, each of which has a series of subsections. Each question concerns one particular class of things. The subsections ask you to estimate how often, in every one thousand occurrences, an item will have a particular feature. So, for example, you may be asked:

Out of every 1000 students living in university accommodation, how many would you expect to

(a) own a car?
(b) have a part-time job?

If you feel that the answer to the first part of the question is 155, then you should write this in the space provided.

Please try to answer all of the questions. While it is unlikely that you will know the exact answer to any of the questions, we are interested in your best guess. This does not mean that we want you to put down the first number that pops into your head. Your knowledge of each of the things in the questions should enable you to give sensible answers in most cases!

If you have any questions, please raise your hand and the experimenter will help you.

On the following pages were a set of questions concerning objects which possessed the feature which had been previously estimated as being rare. The precise form of the questions for each of the three problem contents is given below.

Out of every 1000 houses with a swimming pool, how many would you expect to......;

Out of every 1000 engineers who earn approximately £60,000 a year how many would you expect to.......;

Out of every 1000 cars bought new in 1988 which have a top speed of over 165 mph, how many would you expect to..... .

Although for each set of questions there were only two estimated base rates which were of interest, the questions concerning these base rates were placed in a list of between six and nine other items. This was because of the need to pre-test materials for other experiments. All of these items are presented in Appendix 4. The estimated base rates for some of these other features will be discussed in Chapter 5.
Design: as there were only two features of interest for each problem content, this part of the pre-test had a 2x3 mixed design. The between subjects factor was degree of association (Association) whilst the within subjects factor was problem content.

Procedure: subjects in this part of the pre-test rare correlation condition were run in several groups of between 6 and 10. All subjects were given 15 minutes to complete the booklet.

Results

The mean ratings for each of the features of interest are given in Table 4.7. The ratings produced by subjects in this part of the pre-test were analysed using a 2x3 within subjects Anova. The factors in this analysis were Association\(^1\) (high and low) and

<table>
<thead>
<tr>
<th>HOUSE (with a swimming pool)</th>
<th>ENGINEER (earning £60,000 pa)</th>
<th>CAR (top speed of 165mph+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>garage 94% (9)</td>
<td>drives a company car 84% (16)</td>
<td>radio 95% (18)</td>
</tr>
<tr>
<td>built between 58% (24)</td>
<td>works in general 36% (26)</td>
<td>four doors 50% (26)</td>
</tr>
<tr>
<td>1945-85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7: The rate at which objects possessing the rare feature were expected to possess each of the common features (standard deviations in parentheses) from the pre-test for Experiment 4.

There was a significant main effect for Association F(1, 29) = 239.93, p < .001. As expected, the mean frequency rating for the predicted high association features was higher (mean = 91%) than was the mean rating for the predicted low association features (mean = 48%). There was also a significant main effect of problem content F(2, 58) = 10.06, p < .001. The mean rating for the engineer problem (mean = 60%) was

\(^1\) Association is used here in a restricted sense. It is intended to refer to the case where possession of one feature is associated with possession of a second feature. A more strict, correlational, definition of association would be Association = ABS[Prob(Common feature/Rare feature) - Prob(Common feature)]. As may be seen from a consideration of Tables 4.6 and 4.7, there is very little difference between the High and Low Association conditions when this strict definition is applied. However, the logic of this experiment only requires that subjects in the High Association condition, initially told that a high percentage of instances of X possess the Rare feature have higher expectations concerning the Common feature than do subjects in the Low Association condition. As may be seen from Table 4.7, this is the case. Ideally, Expectation should have been term used to describe this manipulation. However, it has already been used to describe the second dependent variable in Experiment 3 and is thus unavailable.
significantly lower than the mean ratings for the house and car problems (mean ratings of 76% and 72%, respectively). However, the interaction between predicted correlation and content was not significant $F(2, 58) = 1.41, p > .25$). The mean ratings for all of the features used in this part of the pre-test are given in Appendix 4.

One aspects of these results deserves a brief mention. The finding of significant effects of content leads to the expectation that there will be significant effects for content in Experiment 4. In practice, it was impossible to control for these differences.

4.3.3 Experiment 4 - Method

Subjects: 72 subjects took part in this experiment. 25 of these were male and 47 were female. There were two distinct groups of subjects who participated in the experiment: 31 from the first year Business Studies Department at the University of Plymouth; and 41 from the first and third year classes in the Psychology Department at the University of Plymouth. Subjects from these two distinct groups were evenly distributed across conditions. The mean age of the subjects was 23.4. The oldest subject was 43 and the youngest was 17.

Materials: subjects received a handout which comprised of an instruction sheet and three problems. These instructions were as follows:

Accompanying these instructions is a series of three decision problems which require you, in the light of certain information, to rate your confidence in two alternatives and to answer some simple questions. Detailed instructions on what to do are contained in the problems but it would be very helpful if you could bear the following general points in mind throughout the experiment.

1) Read each problem carefully and think hard before responding.
Although there are some surface similarities between the problems, they all concern different scenarios, so you should think carefully about each scenario.

Don't refer back to previous problems when working on the later problems.

Each problem consists of two pages which are stapled together. It is important that you receive the information they contain in the intended order. Don't go on to a later page until you have finished the one you are currently working on.

If you have any questions at any time, please raise your hand and the experimenter will answer them.

The structure and content of the problems was similar to that used in Experiment 3. Thus, for each problem, subjects were given a scenario involving a decision about which of two alternatives was likely to be the case. Next they received a piece of evidence relevant to the decision and were asked to rate their confidence in the alternatives in the light of the evidence they had received. Following this, subjects were asked to state their expectations about the remaining unseen evidence. On the next page was a further piece of information relevant to their decision and a second rating scale. As in Experiment 3, all of the evidence which subjects received was relevant to just one of the alternatives presented in the scenario. Each subject received three problems involving a house, an engineer and a car.

As stated above, subjects were once again asked about their expectations for each possible piece of evidence. This was done in order to keep the conditions in this experiment as similar as possible to those in Experiment 3. Once again, as a manipulation check, only those answers about the piece of evidence which subjects were to receive will be analysed here.
There were four conditions in this experiment. In two of the four conditions subjects received the information in the order of the example given for Experiment 3 (see pgs 153 - 155). In the other two conditions subjects received the evidence in the opposite order. The second manipulation in this experiment was achieved by varying the level of perceived association amongst the features with which subjects were presented. For each of the three problem types there were two sets of features, one set which had a high level of perceived association (as in the example on pgs 153 - 155), and the other which had a low level of perceived association. The strengths of these perceived associations were established in the pre-test (see previous section). These feature pairs are shown in Table 4.8.

From the preceding description of the materials it should be apparent that this experiment is merely an extension of Experiment 3. In fact, one half of this experiment is a replication of Experiment 3.

<table>
<thead>
<tr>
<th>Association</th>
<th>Strength</th>
<th>House</th>
<th>Engineer</th>
<th>Car</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Strong</td>
<td>swimming pool</td>
<td>£60,000+ p.a.</td>
<td>165+mph</td>
</tr>
<tr>
<td></td>
<td>Weak</td>
<td>garage</td>
<td>company car</td>
<td>radio</td>
</tr>
<tr>
<td>Low</td>
<td>Strong</td>
<td>swimming pool</td>
<td>£60,000+ p.a.</td>
<td>165+mph</td>
</tr>
<tr>
<td></td>
<td>Weak</td>
<td>built between</td>
<td>works in general construction</td>
<td>four doors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1945-85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.8: Combinations of high and low association features used in Experiment 4 broken down by whether the feature formed part of a strong or weak piece of evidence.

*Design*: there were two parts to this experiment. The first involved subjects' ratings of their confidence in the hypotheses. This part of the experiment had a 2x2x2x3 mixed design, with the between subjects factors being the perceived association between the features (Association) and the order in which information was presented to subjects (Order), and the within subjects factors being whether subjects rated their confidence in the hypotheses before or after presentation of the second piece of evidence (Before/After) and problem content. The second part of the experiment involved subjects' expectations about the second piece of evidence which they received. As this measure was taken prior to receipt of the second piece of evidence, the Before/After factor is not involved here. This
measure therefore involves a 2x2x3 design. The design of both parts of the experiment is summarised in Table 4.9.

<table>
<thead>
<tr>
<th>Task</th>
<th>Independent Variables</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Ratings</td>
<td>1. ASSOCIATION</td>
<td>2 - HIGH, LOW</td>
</tr>
<tr>
<td></td>
<td>2. ORDER</td>
<td>2 - STRONG 1ST; STRONG 2ND</td>
</tr>
<tr>
<td></td>
<td>3. BEFORE/AFTER</td>
<td>2 - Confidence Before Confidence After</td>
</tr>
<tr>
<td></td>
<td>4. PROBLEM CONTENT</td>
<td>3. Engineer, House, Car</td>
</tr>
<tr>
<td>Expectations</td>
<td>1, 2 and 4</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9: Independent and dependent variables from Experiment 4.

Procedure: as this experiment was run in parallel with several others, two thirds of the subjects were run as part of two groups of one hundred. The other third of the subjects were run as part of several smaller groups of between ten and sixteen subjects. All subjects participated in the experiment during class-time. Each of the four conditions contained approximately equal numbers of subjects from both sizes of groups. In each condition three subjects received the problems in each of the six possible orders. They worked on their problems individually.

4.3.4 Results

4.3.4.1 Confidence ratings

Subjects' confidence ratings were analysed using a 2x2x2x3 mixed design analysis of variance. The between subjects factors were Association and Order, whilst the within subjects factors were Before/After and problem content. The means and standard deviations from this analysis are given in Table 4.10, whilst a full Anova table is given in Appendix 2. Once again it will be remembered that subjects were asked to express their confidence on a 100mm line. At one end of the line was labelled "certain...X" and at the other end "certain...Y". For the purposes of the analysis which follows subjects' markings on the line were converted to scores on a 100 point scale ranging from 1 to 100. Each point
on this scale corresponds to one millimetre on the line and the higher is a subject's score on this scale the more confident she is that the X hypothesis is the case.

The main effect for Before/After was highly significant $F(1, 68) = 22.02, p < .001$. The mean confidence before rating was 66 whereas the mean confidence after rating was 73. There were no other significant main effects. The interaction between Order and

<table>
<thead>
<tr>
<th>Associat'n</th>
<th>Order</th>
<th>Before</th>
<th></th>
<th>After</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>House</td>
<td>Engineer</td>
<td>Car</td>
<td>House</td>
</tr>
<tr>
<td>High</td>
<td>Strong 1st</td>
<td>68</td>
<td>16</td>
<td>66</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>63</td>
<td>16</td>
<td>63</td>
<td>16</td>
</tr>
<tr>
<td>Low</td>
<td>Strong 1st</td>
<td>65</td>
<td>20</td>
<td>69</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>64</td>
<td>13</td>
<td>65</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 4.10: Means (in bold) and standard deviations from the analysis of subjects' confidence ratings in Experiment 4.

Before/After was significant $F(1, 68) = 11.57, p < .002$. This interaction is shown in Figure 4.4, whilst the means and standard deviations involved in the interaction are given in Table 4.11. Tukey HSDs revealed that neither the difference between the means produced before receipt of the second piece of evidence, nor the difference between the means produced after, was significant ($p > .05$ in both cases). Thus, the overall recency effect in this experiment was not significant.

![Figure 4.4: Interaction of Order with Before/After from the analysis of subjects' confidence ratings in Experiment 4.](image)
Table 4.11: Means and standard deviations (in parentheses) involved in the significant interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 4.

<table>
<thead>
<tr>
<th>Association</th>
<th>Order</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Strong 1st</td>
<td>69 (16)</td>
<td>69 (14)</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>63 (15)</td>
<td>79 (16)</td>
</tr>
<tr>
<td>Low</td>
<td>Strong 1st</td>
<td>68 (20)</td>
<td>72 (17)</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>64 (9)</td>
<td>71 (16)</td>
</tr>
</tbody>
</table>

Table 4.12: The means and standard deviations (in parentheses) involved in the significant interaction between Association, Order and Before/After from the analysis of subjects' confidence ratings in Experiment 4.

Tests for simple interaction effects on the significant three way interaction revealed a significant interaction between Order and Before/After when the perceived association between the features was high (F(1, 68) = 5.09, p < .05), but not when the perceived association between the features was low (F(1, 68) < 1). Tukey HSDs were performed on the significant half of the three way interaction. Those subjects who received the strong piece of evidence second produced a mean confidence rating after receipt of the strong evidence which was significantly higher than the remaining three means involved in the interaction. The increase in confidence, due to receipt of the second piece of evidence, for this group was highly significant (p < .0002), whilst both of the other differences were significant at p < .02. None of the other differences between the means involved in this interaction were significant.
4.3.4.2 Expectations about the second piece of evidence

A total of 16 subjects failed to respond to at least one of the questions measuring expectation so they have been excluded for the purposes of this analysis. Of the remaining 56 subjects there were 13 in the high association/strong information first condition; 15 in the high association/weak information first condition, and 14 in each of the remaining conditions. A 2x2x3 mixed design Anova was used to analyse results on this measure. The between subjects factors were Association and Order, whilst the within subjects factor was problem content. The means and standard deviations from this analysis are given in Table 4.13, whilst a full Anova table is given in Appendix 2.

<table>
<thead>
<tr>
<th>Association</th>
<th>Order</th>
<th>House</th>
<th>Engineer</th>
<th>Car</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Strong 1st</td>
<td>86 17</td>
<td>80 19</td>
<td>87 17</td>
<td>84 16</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>30 27</td>
<td>48 30</td>
<td>48 36</td>
<td>42 25</td>
</tr>
<tr>
<td>Low</td>
<td>Strong 1st</td>
<td>69 25</td>
<td>63 30</td>
<td>61 31</td>
<td>64 22</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>33 27</td>
<td>26 25</td>
<td>46 31</td>
<td>35 24</td>
</tr>
</tbody>
</table>

Table 4.13: Means (in bold) and standard deviations from the analysis of subjects' expectations in Experiment 4.
There was a significant main effect for Association $F(1, 52) = 5.22, \ p < .03$; subjects in the high association conditions had a mean expectation of 63 whilst those in the low association condition had a mean expectation of 50. There was also a significant main effect for Order $F(1, 52) = 36.80, \ p < .0001$; subjects who received the strong information first had a mean expectation of 74 whilst those who received the weak information first had a mean expectation of 39. The interaction between these two factors was not significant ($F(1, 52) = 1.26, \ p > .25$)

There was a significant two-way interaction between Order and problem content $F(2, 104) = 3.34, \ p < .04$. Tukey HSD tests revealed significant differences amongst the means involved in this interaction. These means and their standard deviations are shown in Table 4.14, whilst the interaction is represented in Figure 4.6. The differences between means were, with just one exception, as expected. Each mean expectation produced by subjects who received the strong evidence first, was significantly higher than each mean produced by those subjects who received the strong evidence second ($p < .001$ in all cases). The one unexpected result was the significant difference in the strong information 2nd condition between mean expectations for the second piece of evidence in the house problem and mean expectations in the car problem ($p < .05$). None of the other interactions involved in this analysis were significant.

![Figure 4.6](image.png)

Figure 4.6: The significant interaction between Order and problem content from the analysis of subjects' expectations in Experiment 4.
Table 4.14: Means and standard deviations (in parentheses) involved in the interaction between Order and problem content from the analysis of subjects' expectations in Experiment 4.

<table>
<thead>
<tr>
<th></th>
<th>House</th>
<th>Engineer</th>
<th>Car</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong 1st</td>
<td>77 (23)</td>
<td>71 (26)</td>
<td>73 (28)</td>
</tr>
<tr>
<td>Strong 2nd</td>
<td>32 (26)</td>
<td>38 (30)</td>
<td>47 (33)</td>
</tr>
</tbody>
</table>

4.3.5 Discussion

Both of the predictions made at the outset of this experiment have been confirmed. Firstly, the features which were used in Experiment 3 were shown to have a high degree of perceived association. By this it is meant that subjects expected a very high proportion of instances of the objects used, which possessed the rare feature, to also possess the common feature. Secondly, although the recency effect of Experiment 3 was replicated when subjects were given information concerning these associated features, no recency effect was obtained when the features used were not perceived by subjects to be associated. As was stated in the introduction to this experiment, such a result strongly supports an account of the belief revision process which implicates expectation, and suggests that the anchor and adjust model of Hogarth and Einhorn is an insufficient account of the belief revision process.

Subjects' expectations about the unseen evidence are also of interest. The significant main effects of the Association and Order manipulations suggest that these manipulations were successful. Subjects, who received initial information that the majority of instances of an hypothesised category possessed a certain feature, had significantly greater expectations concerning the second feature when that feature was perceived to be associated with the first feature. Likewise, subjects, given initial information that the majority of instances of a hypothesised category possessed a very rare feature, expected the majority of those instances to also possess the common feature. On the other hand, subjects, when told that the majority of instances of the hypothesised category possessed a common feature, had significantly lower expectations concerning the level at which the rare feature was likely to be present.
The two way interaction between Association and Order, although insignificant, was also very interesting (see Figure 4.13). Whilst initial information about a rare feature produced high expectations concerning both types of common feature, these expectations were higher for the associated common features than for the non-associated features. The equivalent difference between the expectations of subjects given an initial piece of information about a common feature was very slight and suggests that differences in expectations did not play a major role in the belief revision of those subjects who received initial information about the rare feature.

Although a detailed discussion of the implications of these results for theories of belief revision will be left until the general discussion, the performance of those subjects who received information about non-associated features will be briefly discussed here. What is most interesting about the results of this half of the experiment is the complete disappearance of the recency effect. This disappearance is probably the cause of the non-significance of the overall recency effect in the experiment. The complete absence of a recency effect may also be taken as evidence of rational information use on the part of subjects. Although the information which subjects received was relevant to only one of the hypotheses and therefore, according to the normative analysis adopted by workers such as Mynatt, Doherty and Dragan (1993), non-diagnostic, subjects did seem to use that information in a manner which would be expected. Those subjects who were given the strong information first had a higher level of belief in the focal hypothesis after receipt of this first piece of information than had subjects who received the strong piece of evidence second. However, the second piece of information had a greater effect on subjects' initial beliefs when that second piece of information was the stronger one. In fact, the effect of a strong second piece of evidence was to raise the confidence of subjects to the same level as that of subjects who received the strong piece of evidence first. What is surprising is the fact that the final confidence ratings of all of the subjects who received information concerning features which had low perceived association were lower than the final confidence ratings of subjects who received information about high association features in the weak/strong order. This is surprising because expectations were not predicted to play a
significant effect in the confidence ratings of any of these three groups of subjects. As both the perceived frequency of the features about which these subjects received information and the rate at which these features were said to be possessed by instances of category X were equal, the final confidence ratings of all three groups of subjects should also have been approximately equal. Although Figure 4.5 seems to suggest that there is a large difference between the final confidence ratings of those subjects in the high association/strong 2nd condition and the final ratings of subjects in both of the low association conditions, this difference was not found to be significant.

Of course, because the evidence which subjects received in this experiment was "non-diagnostic", according to the traditional normative analysis of the pseudo-diagnosticity task (e.g. Mynatt, Doherty and Dragan, 1993), subjects in both halves of the experiment should have expressed equal confidence in the X and Y hypotheses after receipt of each piece of evidence. Nevertheless, it may be claimed, on the basis of these results, that the "irrational" effect of information order is due, at least in some cases, to the existence of "rational" (e.g. Dennett, 1991) expectations. In this light, it is worth pointing out that although an alternative normative analysis of information use on the pseudo-diagnosticity task will not be attempted in this thesis, the normative analysis of the task used in Experiment 2 suggests that an alternative normative analysis of information use is also possible. As with the analysis presented in Chapter 2, subjects' expectations about unseen evidence are likely to be crucial in any alternative normative analysis of information use.

4.4 EXPERIMENT 5

4.4.1 Introduction

The major finding of the previous studies in this chapter was that Hogarth and Einhorn's contrast assumption is incapable of predicting when recency will, and will not, occur. It was concluded that this is so because Hogarth and Einhorn's model does not take into account the degree of relationship which may exist between two pieces of evidence. In
Experiment 4, degree of contrast was controlled for by means of a rating study.

Experiment 5 will constitute an attempt to further demonstrate that Hogarth and Einhorn's contrast assumption is an inadequate predictor of recency. As was the case in Experiment 4, degree of contrast will be controlled for in Experiment 5. In this experiment, however, degree of contrast will be controlled for by using the same features in each half of the experiment. This is best illustrated by the example which follows.

Consider the subject who is asked to decide which model of car her sister owns. She is initially told that 95% of car Xs possess a top speed of over 165 mph. Next she is told that 70% of car Xs possess a radio. In Experiments 3 and 4 these types of materials, in the above order, produced roughly identical confidence ratings before and after receipt of the second piece of evidence. When the order in which subjects received the evidence was reversed, a significant increase in ratings after receipt of the second piece of evidence was found. Experiment 4 demonstrated that such an effect is due to the degree which possession of the common feature is predicted by possession of the rare feature. When subjects received information about two features which were equally different in terms of their perceived frequency (and therefore likely to have equal "subjective value" for the subject), but where the rare feature was not perceived to predict the common feature, no order effect was observed.

It is possible to imagine a very similar experiment where it is the likelihood that members of category X possess each of the features which is manipulated, rather than the strength of the perceived relationship between the features (in all cases the perceived association between the features would be high). For example, one group of subjects might first be told that 95% of car Xs possessed a top speed of 165 mph and then be told that 95% of car Xs possessed a radio, whilst another group might initially be told that 70% of car Xs possessed a top speed of 165 mph and subsequently be told that 70% of cars Xs possessed a radio. In this case both sets of subjects are receiving information likely to be equally different in subjective value. This is because both groups receive information about the same features, and both are told that these features are possessed by the focal category at the same rate. Although one set of subjects is told that both features are
possessed by 95% of instances of category X, whereas the other set of subjects are told that both features are possessed by 70% of instances of category X, no differences in degree of contrast between groups should be caused by this manipulation.

Due to the contrast assumption, Hogarth and Einhorn's model of belief revision would predict a similar sized recency effect for both groups of subjects. However, an account of belief revision which implicated expectation would not make such a prediction. It would claim that a recency effect was much more likely in the case where subjects were told that 95% of instances of the focal category possessed both of the evidential features. This is because of the predicted difference between subjects' expectations about the second piece of evidence which they will receive and the actual value of that evidence. Subjects told that 95% of instances of car X possessed a top speed of over 165 mph will have higher expectations about the second piece of evidence than will subjects told that 70% of car Xs possess such a top speed. However, in both cases it is likely that those expectations will be met by the evidence which subjects receive. Subjects told that 95% of instances of the focal (X) category possess the rare feature will also be told that 95% of instances of that category possess the common feature. Likewise, subjects told that 70% of instances of the focal category possess the rare feature will also be told that 70% of instances possess the common feature. In both cases, the information which subjects receive would be predicted to meet subjects' expectations about that information. Accordingly, these subjects would be predicted to revise their beliefs upwards to the same extent.

However, the final confidence ratings of those subjects who receive an initial piece of information about a common feature are expected to differ significantly. If expectation is implicated in the process of belief revision, then the final confidence ratings of subjects initially told that 95% of instances of the focal category possess the common feature are expected to be significantly higher than the final confidence ratings of subjects told that 70% of instances of the focal category possess the common feature. Once again this prediction is made based upon the predicted differences between these subjects' expectations about the unseen evidence and the actual value of that evidence. Consider the subject told that 95% of instances of the focal category possess a common feature. She will
expect a relatively low proportion of instances of that category to possess the rare feature. However, the subsequent information that 95% of instances of the focal category possess the rare feature will greatly exceed her expectations thereby causing her to revise her initial confidence rating upwards in favour of the focal hypothesis. Now consider a second subject, told that 70% of instances of the focal category possess the common feature. Her expectations about the unseen evidence would be predicted to be approximately equal to those of the earlier subject. This is because although predictions may be made about possession of the common feature given possession of the rare feature, the reverse is not the case. Accordingly, changes in the likelihood that instances of the focal category possess the common feature should not dramatically affect subjects' predictions about the likelihood that instances of that category will possess the rare feature. This second subject will be told that 70% of instances of the focal category possess the rare feature. Although this piece of evidence is predicted to be stronger than subjects' expectations, it will not exceed those expectations to the same extent that being told that 95% of instances possess the rare feature would. For this reason, it is predicted that the final confidence rating of this subject in the 70% condition would be lower than the final confidence rating of the previous subject in the 95% condition.

There are two possible outcomes given an experiment such as the one just described. Firstly, it may be the case that recency will be observed in both halves of the experiment. However, even if this is the case, a significantly larger recency effect is predicted amongst subjects told that 95% of instances of category X possess both of the features. This is because subjects in this condition, who receive the stronger piece of evidence second, are predicted to revise their initial confidence ratings upwards to a greater degree than will the equivalent subjects in the 70% condition. The second possible outcome is that of a recency effect in the 95% condition and no such effect in the 70% condition. It may be the case that in order for subjects who receive the strong piece of evidence second to revise their initial ratings upwards to a significant extent, the final piece of evidence which they receive must dramatically exceed their expectations as it did in Experiments 1, 3 and 4. Either of these outcomes would constitute evidence against
Hogarth and Einhorn's contrast assumption however which predicts equal recency in both halves of the experiment.

Accordingly, the experiment which follows may be thought of as having two purposes. Firstly, it seeks to further examine the order effect established in Chapter 4. The specific question which it asks is the extent to which the difference in the likelihood of occurrence attributed to the evidential features in Experiments 3 and 4 contributes to the order effect found in both those experiments. Secondly, the experiment seeks to provide Hogarth and Einhorn's contrast assumption with a further test. It will reduce but at the same time, keep constant, the contrast between the pieces of evidence which subjects receive. This reduction in contrast will be achieved by increasing, or decreasing, the overall strength of the evidence which subjects receive. It is predicted that if a recency effect is found, it will only be found, or be significantly larger, in those conditions where the overall strength of the evidence has been increased.

4.4.2 Method

Subjects: 72 subjects took part in this experiment. 31 of these were male and 41 were female. Subjects were drawn from the same groups as in the previous experiment and were evenly distributed across conditions. Thirty business students and forty two psychology students participated. The mean age of subjects was 23.6. The oldest subject was 47 and the youngest was 18.

Materials: as in Experiment 4, subjects received a handout which comprised an instruction sheet and three problems. The instructions were exactly the same as those used in Experiment 4. The structure and content of the problems was also exactly the same as that used in Experiment 4. Thus, for each problem, subjects were given a scenario involving a decision about which of two alternatives was likely to be the case. Next they received a piece of evidence relevant to the decision and were asked to rate their confidence in the alternatives in the light of the evidence they had received. Following this, subjects were asked to state their expectations about the remaining unseen evidence.
On the next page was a further piece of information relevant to their decision and a second rating scale. As in Experiments 3 and 4, all of the evidence which subjects received was relevant to just one of the alternatives presented in the scenario. Each subject received three problems involving a house, an engineer and a car. The features used were the same as those used in the high association problems of Experiment 4.

As stated above, subjects were once again asked about their expectations for each possible piece of evidence. This was done in order to keep the conditions in this experiment as similar as possible to those in Experiment 3. Once again, as a manipulation check, only those answers about the piece of evidence which subjects were to receive will be analysed here.

As with Experiments 3 and 4, the order in which subjects received pieces of evidence was manipulated. Thus, subjects received pieces of evidence in a strong/weak or weak/strong order. The second manipulation in this experiment was the likelihood that members of the category X possessed each of the evidential features. In Experiments 3 and 4, the common feature was possessed by 70% of instances of category X, whilst the rare feature was possessed by 95% of instances. Half of the subjects in this experiment were told that both features were possessed by 95% of instances of category X, whilst the other half were told that both features were possessed by 70% of instances. This meant that the information order manipulation was achieved simply by varying the order in which subjects received information about the rare and common features.

**Design:** this was a 2x2x2x3 mixed design, with the between subjects factors being the likelihood that the evidential features were possessed by instances of the focal (X) category (Likelihood), and the order in which information was presented to subjects (Order). The within subjects factors were when subjects rated their confidence in the hypothesis (Before/After), and problem content.

As with Experiment 4, the second measure was subjects' expectations about the second piece of evidence. As this measure was taken before receipt of the second piece of
evidence, the Before/After factor is not involved here. This measure therefore had a 2x2x3 design. The design of both parts of this experiment is summarised in Table 4.15.

<table>
<thead>
<tr>
<th>Task</th>
<th>Independent Variables</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Ratings</td>
<td>1. LIKELIHOOD</td>
<td>2 - 95%, 70%</td>
</tr>
<tr>
<td></td>
<td>2. ORDER</td>
<td>2 - STRONG 1ST; STRONG 2ND</td>
</tr>
<tr>
<td></td>
<td>3. BEFORE/AFTER</td>
<td>2 - Confidence Before Confidence After</td>
</tr>
<tr>
<td></td>
<td>4. PROBLEM CONTENT</td>
<td>3. Engineer, House, Car</td>
</tr>
</tbody>
</table>

Table 4.15: Independent and dependent variables from Experiment 5.

Procedure: as this experiment was run in parallel with several others, two thirds of the subjects were run as part of two groups of one hundred. The other third of the subjects were run as part of several smaller groups of between ten and sixteen subjects. All subjects participated in the experiment during class-time. Each of the four conditions contained approximately equal numbers of subjects from both sizes of groups. In each condition three subjects received the problems in each of the six possible orders. They worked on their problems individually.

4.4.3 Results

4.4.3.1 Confidence measures

In this section the analysis of subjects' confidence ratings will be presented. As in Experiments 3 and 4 subjects were asked to express their confidence on a 100mm line. At one end of the line was labelled "certain...X" and at the other end "certain...Y". For the purposes of the analysis which follows subjects' markings on the line were converted to scores on a 100 point scale ranging from 1 to 100. Each point on this scale corresponds to one millimetre on the line and the higher is a subject's score on this scale the more confident she is that the X hypothesis is the case.
One subject in the 70%/strong information first condition failed to complete at least one of the confidence scales. For the purposes of this analysis this subject has been excluded. A 2x2x2x3 mixed design Anova was used to analyse results on the confidence measure. The between subjects factors were Likelihood (95% vs. 70%) and Order. The within subjects factors were Before/After and problem content. The means and standard deviations from this analysis are presented in Table 4.16, whilst a full Anova table is given in Appendix 3.

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Order</th>
<th>Before</th>
<th></th>
<th>After</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>House</td>
<td>Engineer</td>
<td>Car</td>
<td>House</td>
<td>Engineer</td>
</tr>
<tr>
<td>95%</td>
<td>Strong 1st</td>
<td>69 16</td>
<td>66 18</td>
<td>65 17</td>
<td>72 19</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>64 15</td>
<td>63 17</td>
<td>64 16</td>
<td>79 18</td>
</tr>
<tr>
<td>70%</td>
<td>Strong 1st</td>
<td>70 15</td>
<td>64 12</td>
<td>62 20</td>
<td>68 16</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>61 14</td>
<td>58 13</td>
<td>58 10</td>
<td>67 16</td>
</tr>
</tbody>
</table>

Table 4.16: Means and standard deviations from the analysis of subjects' confidence ratings in Experiment 5.

There was a highly significant main effect for Before/After (F(1, 67) = 37.48, p < .001). The mean confidence rating before receipt of the second piece of evidence was 64 (S.D. = 13), whilst the mean rating after was 71 (S.D. = 16). None of the other main effects approached significance. As was the case in Experiments 3 and 4, there was also a highly significant two-way interaction between the Order and Before/After factors (F(1, 67) = 37.48, p < .005). This interaction is represented in Figure 4.7. Tukey HSD tests for unequal sample sizes revealed significant differences between means involved in this interaction. The means and their standard deviations, as well as significant differences between means, involved in this interaction are given in Table 4.17.
The two-way interaction between Likelihood and Before/After was marginally significant (F(1, 67) = 3.56, p < .07). This interaction is represented in Figure 4.8.
As there was good reason for believing that there would be significant differences between means, Tukey HSD tests were once again used. The means involved in this interaction and the significant differences between means are given in Table 4.18.

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Bef/Aft</th>
<th>1 65 (15)</th>
<th>2 75 (17)</th>
<th>3 62 (10)</th>
<th>4 67 (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>before 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>after 2</td>
<td>.0002</td>
<td></td>
<td>.0002</td>
<td>.0006</td>
</tr>
<tr>
<td>70</td>
<td>before 3</td>
<td></td>
<td></td>
<td>.0002</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>after 4</td>
<td></td>
<td></td>
<td>.0006</td>
<td>.03</td>
</tr>
</tbody>
</table>

Table 4.18: Means and standard deviations (in parentheses) and sig. differences between means for the interaction between Likelihood and Order from the analysis of subjects' confidence ratings from Experiment 5.

The three way interaction between Likelihood, Order, and Before/After also approached significance (F(1, 67) = 3.23, p < .08). This interaction is shown in Figure 4.9.

<table>
<thead>
<tr>
<th>Lik'ld</th>
<th>Order</th>
<th>B/A</th>
<th>1 66 (15)</th>
<th>2 70 (16)</th>
<th>3 63 (14)</th>
<th>4 79 (16)</th>
<th>5 65 (10)</th>
<th>6 69 (14)</th>
<th>7 59 (9)</th>
<th>8 66 (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>str 1st</td>
<td>bef 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aft 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>str 2nd</td>
<td>bef 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aft 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>str 1st</td>
<td>bef 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aft 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>str 2nd</td>
<td>bef 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aft 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.19: Means (in bold), standard deviations (in parentheses) and sig. differences between means for the interaction between Likelihood, Order and Before/After from the analysis of subjects' confidence ratings in Experiment 5 (* indicates p < .0005).

Once again Tukey HSD tests revealed significant differences between means involved in this interaction. Means and significant differences are given in Table 4.19. As may be seen from Figure 4.9 the confidence ratings of subjects told that 95% of instances of category X possess both features display a recency effect whereas there is a trend towards primacy in the ratings of subjects told that 70% of instances of category X possess both features.
examination of Table 4.19 reveals that the former recency effect is significant whereas the latter trend towards primacy is not.

Figure 4.9: The interaction between Likelihood, Order and Before/After from the analysis of subjects' confidence ratings in Experiment 5.

4.4.3.2 Expectation measures

A 2x2x3 mixed design Anova was used to analyse subjects' expectations about the unseen evidence. The between subjects factors were Likelihood and Order, whilst the within subjects factor was problem content. A total of 14 subjects failed to respond to at least one of the questions measuring expectation so they have been excluded for the purposes of this analysis. Of the remaining 58 subjects, there were 14 in both the 95%/strong 1st and 70%/strong 2nd conditions, whilst there were 15 subjects in each of the

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Order</th>
<th>House</th>
<th>Engineer</th>
<th>Car</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>Strong 1st</td>
<td>91 18</td>
<td>84 21</td>
<td>86 20</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>36 36</td>
<td>63 36</td>
<td>63 38</td>
</tr>
<tr>
<td>70%</td>
<td>Strong 1st</td>
<td>87 14</td>
<td>69 20</td>
<td>83 16</td>
</tr>
<tr>
<td></td>
<td>Strong 2nd</td>
<td>39 26</td>
<td>41 23</td>
<td>52 19</td>
</tr>
</tbody>
</table>

Table 4.20: Means (in bold) and standard deviations from the analysis of subjects' confidence ratings in Experiment 5.
other two conditions. The means and standard deviations from this analysis are given in Table 4.20, whilst a full Anova table is given in Appendix 2.

There was a highly significant main effect for Order (F(1, 54) = 50.33, p < .001). Subjects in the strong 1st conditions had a mean expectation of 83 (S.D. = 13) whilst those in the strong 2nd conditions had a mean expectation of 49 (S.D. = 23). The main effect for Likelihood also approached significance (F(1, 54) = 3.26, p < .08). Subjects in the 95% likelihood conditions had a mean expectation of 70 (S.D. = 27) whilst those in the 70% likelihood condition gave a mean response of 62 (S.D. = 23).

![Figure 4.10: The interaction between Order and problem content from the analysis of subjects' expectations in Experiment 5.](image)

<table>
<thead>
<tr>
<th>Info</th>
<th>Content</th>
<th>1 89 (16)</th>
<th>2 76 (21)</th>
<th>3 84 (18)</th>
<th>4 37 (31)</th>
<th>5 52 (32)</th>
<th>6 57 (31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>strong</td>
<td>house 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>eng 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>car 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weak</td>
<td>house 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>eng 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>car 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.21: Means, standard deviations (in parentheses) and sig. differences between means for the interaction of Order with problem content from the analysis of subjects' expectations in Experiment 5 (* indicates p < .001).
There was also a significant two-way interaction between Order and problem content \( (F(2, 54) = 7.13, p < .002) \). This interaction is shown in Figure 4.10. Tukey HSD tests for unequal sample sizes revealed significant differences between means. The means involved in the interaction and the significant differences between means, are given in Table 4.21. The interaction between Likelihood and Order was not significant \( (F(1, 54) < .1) \). The means involved in this interaction may be seen in Table 4.20.

4.4.4 Discussion

The results of Experiment 5 are striking and confirm the general predictions made at the outset. As usual there was a significant difference between subjects' confidence before and after receipt of the second piece of evidence. More interesting however, is the absence of a significant overall recency effect. This absence is to be seen in the interaction between Order and Before/After. Although the interaction itself is significant, follow-up tests revealed that the difference between the final ratings of those subjects who received the strong information first and those who received the strong information second was not significant. This finding is interesting because it is in contrast to the results of Experiments 3 and 4 and suggests that the reduction in contrast between the pieces of evidence which subjects received had a significant overall impact on the recency effect.

However, such a conclusion, based on the significant two-way interaction, must be tempered somewhat when the three-way interaction between Order, Before/After, and Likelihood (see Figure 4.9) is examined. Although the interaction itself was not significant, it was felt that the predictions made at the outset were strong enough to warrant the use of follow-up tests to further examine the interaction. These follow-ups revealed that, as predicted, there was a significant recency effect produced by those subjects told that both evidential features were possessed by 95% of instances of the focal category. On the other hand, when subjects were told that the same features were possessed by only 70% of instances of the focal category, a non-significant trend towards primacy is observed. Such a finding confirms the prediction made at the outset of this experiment. Although the degree of contrast between the subjective values of the individual pieces of evidence was
equal in both halves of this experiment, recency was found only when the overall strength of the evidence was raised. Such a finding is not to be explained in terms of raising or lowering the strength of the evidence. Rather, it is the product of the expectations caused by the initial piece of evidence.

Also of interest in the three way interaction is the degree to which subjects revised upwards their confidence in the focal hypothesis upon receipt of the second piece of evidence. In all cases subjects revised their beliefs upwards. This is in contrast to the results of Experiments 3 and 4 where those subjects who received the strong information first displayed either reduced, or unchanged, confidence following receipt of the second piece of evidence. It would seem that in both halves of this experiment, the second piece of information which subjects received went some way towards meeting the expectations of those subjects who received the strong information first. Interestingly, the pattern of subjects' upwards revisions was as suggested in the introduction to this experiment. As predicted, the greatest revision occurred amongst subjects in the 95% condition who received the strong evidence second. Also as predicted, the extent to which subjects who received the strong evidence first revised their beliefs in the focal hypothesis upon receipt of the second piece of information, was the same in both of the Likelihood conditions. Subjects in the 70% condition who received the strong evidence second revised their confidence ratings to a greater degree than did subjects in either of the Likelihood conditions who received the strong piece of evidence first. Accordingly, it may be inferred that the difference between the expectations of subjects in the 70% condition who received the strong evidence second and the actual value of the evidence given to these subjects was not great enough to cause a significant upwards revision of belief.

The two way interaction between Likelihood and Before/After is also interesting. Although this interaction is significant, the results here are that across information orders, subjects in both likelihood conditions were significantly more confident that the focal hypothesis was the case after receipt of the second piece of information than they were before receipt of that information. The significance of the interaction was caused by the increase in confidence amongst subjects in the 95% condition following receipt of the
second piece of evidence being greater than the increase amongst subjects in the 70% condition. This finding further supports the suggestion that the second piece of evidence which subjects received in this experiment went further, in all cases, towards meeting initial expectations than the equivalent evidence in the previous experiments. The significant difference between the final confidence of those subjects told that both features were possessed by 95% of instances of the focal category and that of subjects told that the same features were possessed by only 70% of instances, is as expected.

Also as expected are the results of the manipulation check on subjects' own expectations. Subjects who received initial information about the rare feature had significantly greater expectations than did subjects who received initial information about the common feature. Likewise, the main effect for likelihood also approached significance. Those subjects initially told that 95% of instances of the focal category possessed one of the evidential features expected more instances of that category to possess the remaining feature than did subjects told that the likelihood of the first feature being possessed by instances of the focal category was 70%.

In general, therefore, the main thrust of the findings of Experiment 5 is, once again, to suggest that it is impossible to explain the effect, or integration, of evidence outside of the context in which that evidence is received. The results of this experiment demonstrate that it is not the difference in the subjective values of pieces of evidence which are predictive of the recency effect found in Experiment 3. Rather, it is the relationship between the pieces of evidence, which itself is determined by inferences from background knowledge, which underlies the effect of information order observed here. The experiments to be presented in the next chapter will seek to investigate more closely the effects of background knowledge on the belief revision process. Specifically, they will seek to determine how subjects use background knowledge about both the base rates of evidential features and the meaning of likelihood expressions in order to assign a subjective value to a piece of evidence. Once again, it will be argued that there are very important pragmatic factors at work in this process.
This chapter has provided evidence concerning the nature of the belief revision process. It will be argued that in the light of this evidence, current conceptions of how that process works will have to be altered. Experiment 3 demonstrated the effect of information order on subjects' integration of evidence concerning a hypothesis which was to be evaluated. This order effect was suggested by the results of Experiment 1. The finding in Experiment 4, that the order effect vanished when subjects were given pieces of evidence which did not lead to the formation of expectations, provided concrete evidence that expectation is in some way implicated in the belief revision process. The results of Experiment 5 also support this view. Although the contrast between the pieces of evidence which subjects received was controlled for in Experiment 5, recency was only found when subjects received a final piece of evidence which greatly exceeded the expectations caused by the initial piece of evidence.

What then are the implications of these results for current accounts of the belief revision process? The answer is very simple. The combined results of Experiments 1, 3, 4 and 5 suggest that current models of the process are wholly inadequate. It was argued in Chapter 3 that current models would prove to be inadequate due to a range of theoretical and experimental work which has been carried out over the last ten years. Much of this work comes from the literature on pragmatics (e.g. Shannon, 1988) and concepts (e.g. Murphy and Medin, 1985; Wattenmaker, 1995). The assumption shared by all current models of belief revision which sits most uneasily with this work is that of independent subjective values for each piece of evidence.

The fact that expectation has been implicated in the belief revision process suggests that independent subjective values for pieces of evidence do not exist. Specifically, the results of the experiments reported in this chapter suggest that the subjective value of a piece of evidence is determined on-line, and is thus context-dependent. It has long been recognised in the belief revision literature (Anderson, 1981; Shanteau and Nagy, 1984; Hogarth and Einhorn, 1992) that the subjective value of any piece of evidence must be
sensitive to the context of judgement. Thus, in an impression formation task, the adjective "cold" would be assigned different values depending on whether the subject was asked to judge the target individual's intelligence, likeableness, or fitness for political office. What has rarely been recognised however, is the importance of evidential context in the derivation of the subjective value of any piece of evidence. With the exception of Gestalt psychologists such as Asch (1946), the emphasis in modelling belief revision has been on assuming a fixed subjective value, and accounting for the differential effect of evidence by a process of differential weighting.

The most recent example of such a model is Hogarth and Einhorn's (1992) anchor and adjust model. Much of this chapter has been taken up with consideration of the contrast assumption which is contained in that model. This has been because the anchor and adjust account would seem to be the most plausible which currently exists. The contrast assumption is an admirable attempt to cope with the problem of evidential context. Unfortunately however, it is based on the idea that subjective value may be assumed. Therefore, although the contrast assumption can account for the results of Experiment 3, it cannot account for those of Experiment 4 and 5, where the difference in subjective value between the two pieces of evidence was kept constant but either the degree of association between the evidential features or the strength of the initial piece of evidence were manipulated.

What does it mean to say that the subjective value of a piece of evidence is determined on-line? In a sense, both Lopes (1987) and Hogarth and Einhorn have been claiming that some part of the meaning of some evidence is so determined. Both have recognised the potential for error inherent in the process of deciding for what hypothesis a piece of evidence is marked (i.e. which hypothesis a piece of evidence supports). Neither, however, has given an account of how this happens. Is the context-dependency of meaning best described by Asch as a directional force exerted by early occurring evidence over later occurring evidence? To accept Asch's characterisation of the process would, undoubtedly, be a retrograde step. Instead, the process of assigning a subjective value to a piece of evidence is best described as being inferential in nature.
In Chapter 3 the importance of inference to the process of categorisation was stressed. Specifically, it was suggested that, in the light of work by Wattenmaker (1995) and Asch and Zukier (1984), the process of categorising social phenomena on the basis of their features, involves a great deal of inference. However, it was also pointed out that inference must also be central to the categorisation of object knowledge. How then, might inference be involved in the belief revision task used in the experiments described in this chapter? One very obvious possibility is that when early occurring information leads to high expectations about as yet, unseen, evidence, those expectations form the major premise of a deductive inference. The minor premise is the evidence in the world. This leads to an argument such as the following (in this example, and the next, the expected level is taken from the results of the pre-test):

If my friend's house is on Street X then 94% of houses on Street X will have garages.
70% of houses on Street X have garages.
Therefore, my friend's house is not on Street X

Because this conclusion contradicts the strong piece of information which subjects received initially, they do not adjust their beliefs. Conversely, although those subjects who received the information in the opposite order may be thought of as performing the same type of inference, the premises are not the same. Rather, the major premise will consist of a conditional expressing expectations about the base rate of the rare feature and the minor premise will come from information present in the problem:

If my friend's house is on Street X then 3% of houses on Street X will possess a swimming pool
95% of houses on Street X possess swimming pools
Therefore, my friend's house is on Street X
Because this conclusion agrees with the initial piece of evidence which these subjects received, its effect is to increase confidence in the focal hypothesis.

An advantage of such a characterisation of the process is that it allows for an interpretation of which hypothesis the evidence is marked for (Lopes, 1987; Hogarth and Einhorn, 1992). In addition to this, it provides a framework for the understanding of how evidence comes to have a subjective value. It does so in a manner which not only allows for, but is driven by, both the judgmental and evidential context.

One very interesting feature of the results of Experiment 4 however, is the contrast between the pattern of results for those subjects who received evidence about features where possession of the rare feature was perceived to predict possession of the common feature, and those who did not. Apart from the obvious contrast, those subjects who received information about low association features, in the weak/strong order, might have been predicted to have very low expectations about the second piece of evidence. Indeed, the expectations expressed by these subjects was not significantly different from the expectations of subjects who received evidence about associated features in the weak/strong order. It could be argued that the effect of the second, stronger, piece of information on subjects' beliefs about the hypotheses should have been the same in both cases. Because recency occurred only in that half of the experiment where subjects received information about associated features, it is obvious that the stronger piece of information did not have the same effect in both cases.

One possible explanation for this is provided by Stevenson and Over (1995). These experimenters showed that subjects, when making inferences, are sensitive to the certainty with which they may hold their premises. Thus, Stevenson and Over found that subjects were less confident in conclusions drawn from uncertain premises than they were in conclusions drawn from certain premises. Although there is a similar statistical likelihood that houses which have garages and houses built between 1945 and 1985 will have swimming pools, in the former case the link between the two features is stronger. Accordingly, subjects' confidence in their initial premise may have been greater in the
former case. This in turn would have caused them to have greater confidence in their conclusion, and to have revised their beliefs to a greater extent in the light of that conclusion. What is interesting about this feature of the results is that it illustrates the insufficiency of arguments from probability, in giving a comprehensive account of either belief revision, or how knowledge is organised. In Chapter 3, Murphy and Medin's (1985) argument against accounts based wholly on feature frequency was mentioned. Likewise, in Chapter 1, it was argued that a completely probabilistic account of information selection is most likely to be impossible. Although the studies contained in both this chapter, and Chapter 2, demonstrate the importance of subjective probability to the organism, it is not argued in this thesis that probabilistic information is the sole source of information available to subjects.

This caveat aside, the experiments to be described in the next chapter will further investigate the factors involved in the basic recency effect found in Experiment 3. The effect of the perceived frequency of the individual experimental features used will be examined, as will the effects of background knowledge on the representation of individual pieces of evidence.

4.6 SUMMARY

This chapter has provided evidence supporting the view that the effect of interest in Experiment 1 was due to the sequential nature of the evidence with which subjects were provided. In so doing, these experiments have also provided evidence of the importance of expectations to the process of belief revision. Experiment 3 produced an order effect. Specifically, subjects given strong information following weak information were more confident in the hypothesis to which that information related than were subjects given the same information, but in the reverse order.

Experiment 4 attempted to discriminate between two opposing accounts for the results of Experiment 3. The account of Hogarth and Einhorn (1992) suggests that the recency effect in Experiment 3 occurred because of the difference in the subjective value
of the pieces of evidence which subjects received. An alternative account of the results was presented which claimed that the recency effect was due to the process of deriving the subjective values for the pieces of evidence presented. Such an account would claim that a subjective value is derived by a combination of expectations from background knowledge and information presented in the problem. In Experiment 4, subjects were presented with evidence where the difference in subjective value was controlled for. However, half of the subjects received information about features where the possession of a rare feature was perceived to predict possession of the common feature, and the other half received information about features for which there was no perceived association. The recency effect of Experiment 3 was replicated with the associated features, but no recency was found with the non-associated features.

The results of Experiment 5 also suggest that Hogarth and Einhorn's contrast assumption is an inadequate explanation for the recency effect found in Experiment 3. In Experiment 5 degree of contrast was controlled for by using the same evidential features throughout and telling subjects that instances of the focal category possessed each of the features at the same rate. However, half of the subjects were told that 95% of instances of the focal category possessed both features whilst the remaining subjects were told that 70% of instances of the focal category possessed both features. As predicted, recency was found only amongst the confidence ratings of those subjects told that 95% of instances of the focal category possessed both features.

The results of Experiment 4 and 5 were interpreted as providing evidence for an account of belief revision where the subjective value of a piece of evidence is determined on-line. It was argued that this on-line process is an inferential one, where inferences are made from a combination of information from background knowledge and information present in the problem. It was further argued that such an account of the process explains how evidence comes to be seen as being marked for one hypothesis or the other, and how that same evidence comes to have a subjective value for the subject. Finally, it was pointed out that the results of Experiment 4 suggest that the information drawn on from background knowledge is likely to be both statistical and theoretical.
CHAPTER 5 - EXPERIMENTS 6, 7 AND 8: SOURCES OF EVIDENTIAL STRENGTH

5.1 INTRODUCTION

The previous chapter established the existence of a recency effect in subjects' use of information for the evaluation of hypothesis, and the importance to that effect of both the relationship which exists between those pieces of evidence and the strength of the expectation caused by the initial piece of evidence. In so doing, it suggested that the most recent, and most comprehensive, model of belief revision (Hogarth and Einhorn, 1992) is, at least in some respects, inadequate. This chapter will report the details of three experiments all of which will use the same basic methodology used in Experiments 3, 4 and 5.

To understand the rationale for these experiments it is necessary to, once again, consider Hogarth and Einhorn's contrast assumption. Simply expressed, the contrast assumption states that, in many cases, the occurrence of a recency effect is dependent on the degree to which the pieces of evidence which a subject receives differ in their subjective value. The greater the difference, the more the subject's final expression of belief or confidence will seem to reflect the value of the later evidence she receives. As was discussed in the previous chapter, workers in the area of belief revision use the term subjective value to mean the degree to which the subject understands a piece of evidence to support a particular hypothesis. This begs the question of how the subject comes to assign a subjective value to any piece of evidence. Typically, in the literature on belief revision, some form of pre-test is carried out in order to determine in advance the subjective value of any piece of evidence given to subjects. These pre-tests have ranged from tests of the positivity and negativity of personality adjectives (a technique used by Norman Anderson and his co-workers) to very detailed examinations of individual pieces of evidence (used for example by Tubbs, Gaeth, Levin and van Osdol, 1993).
The materials used in this thesis differ in several respects from those used elsewhere in the literature. Firstly, the pieces of evidence which subjects have received in the course of the experiments described in this thesis have not been pre-tested for their actual subjective value. Rather, various characteristics of the features about which subjects were asked to reason have been pre-tested. Thus it has been claimed that predictions may be made about the direction, and extent, of subjects' belief revision in the face of certain pieces of evidence on the basis of their background beliefs about the base rates of features of objects as well as the extent to which one feature is predicted by another.

A second difference which exists between the materials used in this thesis and, for example, the materials used in personality impression formation tasks, is the slightly more complex nature of the evidence used in the experiments described here. By complexity is not meant the absolute length of the pieces of evidence which subjects receive (absolute length has been used by Hogarth and Einhorn to distinguish simple from complex pieces of information). Instead complexity is intended in this context to refer to the number of specific sources of information available to the subject from background knowledge when she comes to assign a subjective value to a piece of evidence. Consider the subject asked whether her sister's car is a model X or a model Y. She is told:

95% of model Xs have a top speed of over 165 miles per hour
and
70% of model Xs have a radio.

The experiments described in Chapter 4 have demonstrated that information about the extent to which possession of the second feature (having a radio) is predicted by possession of the first feature (having a top speed of over 165 mph) is central to subjects' confidence ratings based on this evidence. However, there are two other sources of information available to subjects in assigning a subjective value to these pieces of evidence. The first of these is their belief about the base rates of the evidential features, whilst the second is their knowledge about likelihood, or probability expressions. How these sources of information might be used by the subjects will be discussed in the
introduction to Experiment 6. Suffice it to say, for the moment, that the experiments described in this chapter will seek to investigate the differential use of these sources of information by subjects in assigning subjective values to pieces of evidence. The first two experiments to be described in this chapter will examine the influence of background knowledge about both the base rates of the features, and the use of likelihood expressions, on the process. Neither of these sources of information from background knowledge have proven to be particularly important to the results of the previous studies. The experiments to be reported here will attempt to demonstrate their use by subjects. By so doing, it is hoped to specify more clearly both the sources of background information upon which the pragmatic inferences, argued in Chapter 4 to be so central to the process of evidence interpretation, are based, and to determine how any one such source assumes more importance than another in a given context. Experiment 8, the final experiment to be described in this thesis, will follow on from the results of Experiments 6 and 7. It will attempt to demonstrate subjects' differential use of information from background knowledge within the same task. In order to do so, an order effects manipulation will once again be used. In the case of Experiment 8, however, subjects will be given information relevant to both hypotheses.

5.2 EXPERIMENT 6

5.2.1 Introduction

The previous three experiments have focused on the order effect suggested by the results of Experiment 1. One of the striking findings of these studies is the importance of the perceived association between features to the occurrence of that effect. It has been argued that the existence of cases where the effect does not occur, due to lack of association between the evidential features used, is evidence that an account of the belief revision process based on Hogarth and Einhorn's contrast assumption, is inadequate. However, it has not been claimed that the contrast assumption will prove to be an inadequate description in every case. Although this claim has not been made, the next
experiment to be described in this chapter will be an attempt to show that even when the contrast assumption does seem to account for the manner in which subjects integrate evidence with hypotheses, no existing model of the process is capable of predicting, a priori, when an order effect will occur. This is because none of these models specify the process by which a subjective value is assigned to a piece of evidence.

The results of Experiments 3, 4, and 5 are best understood as demonstrations of the role played by pragmatic factors in the belief revision process. In summary, it has been demonstrated that early occurring evidence sets up a context in which the meaning of later occurring evidence is determined. As has been pointed out, the process by which the meaning of later occurring evidence is determined, is an inferential one. Current models of the belief revision process cannot account for these results because they lack a pragmatic component. However, as was discussed in Chapter 1, there are two elements to current accounts of pragmatics (e.g. Sperber and Wilson, 1986; 1995). The first of these is pragmatic inference where people infer conclusions from both information in the world and information present in background knowledge. Before a pragmatic inference may be made however, the premises from which such an inference may be made must be selected.

It was argued in Chapter 1 that Sperber and Wilson's Relevance Theory offers an insight into the process by which the information which is important to the organism comes to be attended to. This account is based on what Sperber, Cara, and Girotto (1995) have called "the first (or Cognitive) Principle of Relevance". This has been discussed in some detail in Chapter 1. To recap, the information which is focused upon in any situation is that which has the greatest cognitive effect, and poses the least cognitive effort, for the organism. Thus, being told that the next train to London leaves Plymouth in 2,400 seconds is less relevant to the organism than is being told that it leaves in 40 minutes. This is because an understanding of the former utterance requires more cognitive effort than an understanding of the latter. Likewise, being told that the next train to London leaves at 5.30 p.m. is more relevant than being told that it leaves sometime after 4.00 p.m., because the first piece of information leads to more inferences than the second. Considerations of
effect and effort lead to a trade off between how much work the organism must do to process a piece of information and how useful that information is.

How does such a notion of cognitive effect come into play when making predictions about the task used in the previous three experiments? The answer is relatively simple. There are several sources of information available to subjects engaging in the task. The first of these, the effect of which has already been demonstrated, is knowledge about the extent to which the evidential features are associated. Thus far, all of the experiments examining belief revision in this thesis have used materials where possession of the common feature was predicted by possession of the rare feature. The one exception to this occurred in Experiment 4 where only a weak association existed between the evidential features used. The purpose of the experiments to be described in this chapter is to examine the influence of other sources of information, available from background knowledge, on subjects' interpretation of evidence. Experiment 6 is best characterised as an exploratory study which seeks to investigate the differential use of two other sources of information which are available to subjects.

Before proceeding with a more detailed discussion of these other sources of information, it will be useful to quickly describe the kinds of materials which will be used in Experiment 6. As before, subjects will be asked to express their confidence in two alternatives. So, for example, subjects might be asked to rate their relative confidence that their sister's car is a model X or a model Y in the light of some evidence. Experiment 6 will differ from Experiments 4 and 5 only in terms of the evidence which subjects receive. Half of the subjects will be given information concerning two rare features of the object to be categorised. The remaining subjects will

1. 95% of model Xs have a top speed of over 165 miles per hour
2. 70% of model Xs have leather upholstery.

As may be seen from this example, half of the subjects will be given information concerning two rare features of the object to be categorised. The remaining subjects will
receive evidence concerning two common features of these objects. So, for example, in the context of the car problem these subjects might be told the following:

1. 95% of model Xs have a radio
2. 70% of model Xs have a top speed of over 90 miles per hour.

These features will be pre-tested for their perceived base rates and, as far as is possible, the degree to which any one feature is predicted by the other feature in the pair will be controlled for and maintained at a low level. The order of information manipulation will be achieved by switching the order of the likelihoods. Thus, whilst half of the subjects who receive information about two rare features will receive the information in the first example above, the remaining subjects will receive the following:

1. 70% of model Xs have a top speed of over 165 miles per hour
2. 95% of model Xs have leather upholstery

Likewise, half of the subjects in the common features condition will receive the information in the second example above, whilst the remaining subjects will be told

1. 70% of model Xs have a radio
2. 95% of model Xs have a top speed of over 90 miles per hour.

Thus, a partial order manipulation will be used in this experiment. This will facilitate a clearer understanding of the results of the experiment.

As Experiment 6 is exploratory in nature, no definite predictions may be made about its results. However, some possible results will now be discussed. Firstly, it is possible that a main effect of feature rarity will be found. In other words, it is possible that both the initial, and final, confidence ratings of those subjects who receive information about rare features will be significantly more in favour of the focal hypothesis than the initial, and final, confidence ratings of those subjects who receive information about
common features. This is not considered likely, however, as no such significant differences have emerged in the results of the previous experiments.

It is much more likely that a difference between the confidence ratings of subjects given information about rare features and those given information about common features will be found in the three way interaction between Rarity, Order and Before/After (as before, subjects will be asked to express their relative confidence in the hypotheses both before, and after, receipt of the second piece of information). Owing to notions of cognitive economy it is possible that a recency effect will be found amongst the results of subjects given information about common features but not amongst the results of those subjects given information about rare features. Such a result is possible because of the information which is available to subjects from background knowledge when assigning subjective values to the pieces of evidence which they receive. This background knowledge has to do with firstly, subjects' beliefs about the base rates of the features contained in the evidence and secondly, with subjects' knowledge about the meaning of likelihood expressions. It is possible that this knowledge will be used differentially with common and rare features.

Consider the subject asked to express their relative confidence that their sister's car, which possesses a top speed of over 165 miles per hour and has leather upholstery, is a model X or a model Y. They are told that:

1. 95% of model Xs have a top speed of over 165 miles per hour
2. 70% of model Xs have leather upholstery.

This evidence is highly suggestive that the subject's sister's car is a model X. This is because the features about which the subject has received information are very rare amongst the general population of cars and it is unlikely that model Y cars will possess both of these features at an equally high rate. What is important, or most relevant, about these pieces of information is that both of the (very rare) features are possessed by instances of one of the candidate alternatives with a likelihood that is much higher than
was expected. However, the exact likelihood is unimportant. Subjects' focus, when assigning subjective values to these pieces of evidence, is much more likely to be on the features than on the likelihoods.

Now consider the subject, this time informed that their sister's car has a top speed of over 90 miles per hour and a radio, who is told:

1. 95% of model Xs have a radio
2. 70% of model Xs have a top speed of over 90 miles per hour.

Neither of these pieces of information is particularly suggestive of the subject's sister's car being a model X (although the fact that the experimenter has chosen to give the subject pieces of information relevant to category X might lead the subject to make the pragmatic assumption that she is expected to favour category X). This is because neither of the features are rare in the general population of cars and they are both possessed by model X cars at approximately the rates that would be expected. Therefore, beliefs about the base rates of common features are not as informative (or as relevant) as they are when the features are rare. Thus, in this case it is likely that subjects will focus on the likelihoods to a greater extent than they will in the case where they receive information about rare features.

How might this difference in the information which subjects attend to cause a recency effect when information is received about common features but not when information is received about rare features? If subjects' focus is on the evidential features in the case where they receive information about rare attributes of the object to be classified then there should be very little contrast between the subjective values of the pieces of evidence. Because the features are of equal rarity, and if subjects are focusing on the features, then they should be understood as providing approximately equal support for the hypothesis that the instance to be categorised is a member of category X. On the other hand, if, when they receive information about two common features, subjects are focusing to a greater extent on the likelihoods, then there should be some contrast between the
subjective values which subjects assign to the pieces of evidence. This is because, once again, even though the features about which subjects receive information are equally common, they are focusing upon the likelihoods which do differ (from 95% to 70%). In this case recency should be observed.

Before proceeding with a description of the experiment sketched above, two points must briefly be made. The first is that the predictions discussed above cannot be made by any existing account of the belief revision process. They are made based upon a consideration of the influence of pragmatic factors on the determination of a subjective value for a piece of evidence. Existing accounts of the process, both algebraic (Anderson, 1981), and procedural (Lopes, 1987; Hogarth and Einhorn, 1992), assume a subjective value for any piece of evidence. They do not address themselves to the issue of how evidence is interpreted by the subject. The predictions made here assume that the value of any piece of evidence is context dependent.

The second point which should be made concerns the recency effect predicted with common features in this experiment. Although such an effect is predicted by Hogarth and Einhorn's contrast assumption, it is also predicted by an averaging model of the sort put forward by Anderson (1981), Shanteau and Nagy (1984), and Lopes (1987). The recency prediction made in this instance is not directly motivated by any of these different models. Rather it is motivated by the results in the experimental literature (discussed in Chapter 3) which suggest that averaging-type behaviour does seem to take place with simple stimulus materials. The purpose of the experiment to be described is not to provide support for any of the above-mentioned models. Rather it is to demonstrate their inadequacy in terms of predicting when, and with what stimulus materials, a recency effect will occur.

5.2.2 Method

Subjects: 72 subjects took part in this experiment. Of these 21 were male and 51 were female. 29 were first year undergraduate Business Studies students and 43 were second or third year psychology students. As before all subjects were students at the University of
Plymouth. Subjects from both groups were evenly distributed across conditions. Subjects' mean age was 22.8 years, with the youngest being 18 and the oldest 40.

Materials: as with Experiments 4 and 5, subjects received a handout which comprised an instruction sheet and three problems. The instructions were exactly the same as those used in Experiments 4 and 5. The structure and content of the problems was also exactly the same as that used in Experiment 4 and 5. Thus, for each problem, subjects were given a scenario involving a decision about which of two alternatives was likely to be the case. Next they received a piece of evidence relevant to the decision and were asked to rate their confidence in the alternatives in the light of the evidence they had received. Following this, subjects were asked to state their expectations about the remaining unseen evidence. On the next page was a further piece of information relevant to their decision and a second rating scale. As in Experiments 3 and 4, all of the evidence which subjects received was relevant to just one of the alternatives presented in the scenario. Each subject received three problems involving a house, an engineer and a car. Examples of each of the problem contents used in this experiment are to be found in Appendix 1.

As stated above, subjects were once again asked about their expectations for each possible piece of evidence. This was done in order to keep the conditions in this experiment as similar as possible to those in Experiment 3. Once again, as a manipulation check, only those answers about the piece of evidence which subjects were to receive will be analysed here.

As with Experiments 3, 4 and 5, the order in which subjects received pieces of evidence was manipulated. Thus, subjects received pieces of evidence in a strong/weak or weak/strong order. Unlike previous experiments this manipulation was achieved simply by reversing the order in which subjects received the likelihood information. Thus subjects in the strong/weak condition were told that 95% of instances of X possess the first feature and then told that 70% of instances of X possess the second feature. Conversely, subjects in the weak/strong condition were told that 70% of instances of X possess the first feature and were then told that 95% of instances of X possess the second feature. Most
importantly, in order to achieve the Order manipulation in this experiment the order in which subjects received information about the features was not varied. Thus all subjects received an initial piece of information about the same feature.

The second manipulation in this experiment was the rarity of the evidential features used. In Experiments 4 and 5, subjects received information about one common and one rare feature of the objects they were reasoning about. Half of the subjects in this experiment were given information about two rare features whilst the other half were given information about two common features.

Also like Experiments 4 and 5, there were four conditions in this experiment. These were achieved by crossing the perceived base rates of the feature pairs with information order. Thus, one group of subjects received information concerning two rare features where the first feature was said to be present in 95% of instances of the focal category and the second was said to be present in 70% of instances. For example:

95% of houses on street X possess a swimming pool

and

70% of houses on street X are worth £250,000.

In order to achieve the Order manipulation the remaining subjects received the following information:

70% of houses on street X possess a swimming pool

and

95% of houses on street X are worth £250,000

The remaining two groups of subjects received information concerning common features. Again, one of these groups were told that 95% of instances of the focal category possessed the first feature, whilst the other group were told that 70% of instances possessed the first feature.

As stated earlier, the Rarity manipulation was achieved by giving subjects evidence about either two rare evidential features or two common evidential features.
These features were chosen on the basis of the pre-test described in Chapter 4. One of the features from all of the common and rare pairs had already been used in the previous experiments (see Table 5.1). The other feature in each pair was chosen with reference to both parts of the pre-test. Thus, each feature pair contained features which were perceived to be of approximately equal rarity. An attempt was also made, by consultation with the results of the second part of the pre-test, to control for the strength of perceived association between the features in each pair. As far as possible all of the perceived associations between the features used in this experiment were kept approximately equal. An examination of Table 5.1 reveals that the most obvious exception was the rare feature pair for the engineer scenario. For this pair the perceived association was much weaker than for the other pairs. This was because it was felt that the new feature chosen for the engineer scenario was the one most suitable for problem construction.

Further examination of Table 5.1 reveals two things. Firstly, that subjects' expectations about the percentage of instances already possessing the old Common feature which will also possess the new Common feature are higher than the equivalent expectations for Rare features. Whether this has any effect on the results of this experiment will be very easy to infer. Experiment 5 demonstrated that the extent to which subjects expectations about the second piece of evidence are exceeded by the actual value of that evidence, in some cases, determines the extent to which they revise their beliefs in the light of that evidence. As the expectations of subjects in the Rare conditions are likely to be lower than the expectations of subjects in the Common conditions, if those expectations influence the results of this experiment then we would expect subjects in the Rare conditions to revise their beliefs to a greater extent after receipt of the second piece of evidence than will subjects in the Common conditions.

Further examination of Table 5.1 reveals a second difference between the Rare and Common features. This difference involves the degree of association between the features as measured by the absolute value of $C - B$ (i.e. expectations about the new feature given possession of the old feature - expectations about the new feature). A quick glance at Table 5.1 shows that the Rare features are more closely associated on this measure (mean = 36)
than are the Common features (mean = 5). It is not clear how this difference might affect the results of the study. In general, this measure of association may determine the confidence with which subjects would make predictions about the unseen evidence. A high degree of such association does not imply, given a high rate of possession of the old feature, a high rate of the new feature. Instead, it is a measure of how rate of possession of the new feature is affected by possession of the old feature. Given, however, that possession of the new Rare feature seems to be positively affected by possession of the old Rare feature (more so than is the case for the Common features), it may be the case that subjects in the Rare conditions will have stronger expectations concerning the second piece of evidence than subjects in the Common conditions. If these expectations are implicated in the results of the experiment then, by the same argument as was used in the previous paragraph, there should be less belief revision in the Rare conditions.

It should also be borne in mind that the pre-test only produced data about expectations concerning the second (new) feature of each pair given the presence of the features which have been used throughout this series of experiments (the old features). However, as the order in which subjects received the information about the features was not manipulated in this experiment this data will suffice for an interpretation of the results.

The order of information manipulation was achieved as it had been in the earlier

<table>
<thead>
<tr>
<th>Rarity</th>
<th>Prob</th>
<th>Old feature</th>
<th>A</th>
<th>New feature</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>com</td>
<td>house</td>
<td>garage</td>
<td>56%</td>
<td>garden</td>
<td>81%</td>
<td>82%</td>
</tr>
<tr>
<td></td>
<td>engineer</td>
<td>company car</td>
<td>62%</td>
<td>earns £25,000 p.a.</td>
<td>49%</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>car</td>
<td>radio</td>
<td>92%</td>
<td>90 mph + top speed</td>
<td>78%</td>
<td>82%</td>
</tr>
</tbody>
</table>

| rare   | house| swimming pool | 3% | worth £250,000            | 8% | 62%|
|        | engineer| earns £60,000 p.a. | 12%| travels abroad regularly | 11%| 25%|
|        | car   | 165 mph + top speed | 6% | leather upholstery       | 13%| 52%|

Table 5.1: Perceived rarity and associations for the features used in Experiment 6. Column A contains the % of instances expected to possess the old features. Column B contains the same information for the new features. Column C contains the % of instances which possess the old feature, also expected to possess the new feature.
experiments. As the level of each feature said to be present in each hypothesised set was either 95% or 70%, and the features were of approximately equal diagnosticity, the strong piece of evidence always concerned the feature present at a level of 95%. However, unlike the earlier experiments, the order in which subjects received information about each feature was kept constant. The first piece of information which subjects received always concerned an old feature. The order of information manipulation was achieved by varying the likelihood of this initial feature.

*Design:* once again this was a 2x2x2x3 mixed design with the between subjects manipulations being rarity of feature pairs (Rarity) and Order. In descriptions of previous experiments, the levels of the Order manipulation have been referred to as Strong 1st and Strong 2nd. For the purposes of this experiment, they will be referred to as 95% - 70% and 70% - 95%. This is because these terms are a more accurate description of the order manipulation used in this experiment. The within subjects manipulations were when the confidence measure was taken (Before/After) and problem content.

The second measure used in this experiment was, again, subjects' expectations about the second piece of evidence. This measure had a 2x2x3 mixed design. The between subjects factors were as with the confidence measure, whilst the only within subjects factor was problem content. The design of both parts of the experiment is summarised in Table 5.2

<table>
<thead>
<tr>
<th>Task</th>
<th>Independent Variables</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratings</td>
<td>1. ORDER</td>
<td>2 - STRONG 1ST; STRONG 2ND</td>
</tr>
<tr>
<td></td>
<td>2. RARITY</td>
<td>2 - Rare features Common features</td>
</tr>
<tr>
<td></td>
<td>3. BEFORE/AFTER</td>
<td>2 - Confidence Before Confidence After</td>
</tr>
<tr>
<td></td>
<td>4. PROBLEM CONTENT</td>
<td>3. Engineer, House, Hotel</td>
</tr>
</tbody>
</table>

Table 5.2: The independent and dependent variables involved in Experiment 6.
Procedure: as this experiment was run in parallel with several others, two thirds of the subjects were run as part of two groups of one hundred. The other third of the subjects were run as part of several smaller groups of between ten and sixteen subjects. All subjects participated in the experiment during class-time. Each of the four conditions contained approximately equal numbers of subjects from both sizes of groups. In each condition three subjects received the problems in each of the six possible orders. They worked on their problems individually.

5.2.3 Results

5.2.3.1 Confidence measures

As with previous experiments subjects were asked to express their confidence on a 100mm line. At one end of the line was labelled "certain...X" and at the other end "certain...Y". For the purposes of the analysis which follows subjects' markings on the line were converted to scores on a 100 point scale ranging from 1 to 100. Each point on this scale corresponds to one millemetre on the line and the higher is a subject's score on this scale the more confident she is that the X hypothesis is the case. Two subjects (one in each of the rare features conditions) failed to complete at least one of the confidence measures. For the purposes of this analysis they have been excluded. A 2x2x2x3 mixed design Anova was used to analyse results on this measure. The between subjects factors were Rarity and Order. The within subjects factors were Before/After and problem content. Both the means and standard deviations from this analysis are given in Table 5.3, whilst a full Anova table is given in Appendix 3.

<table>
<thead>
<tr>
<th>Rarity</th>
<th>Order</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>House</td>
<td>Engineer</td>
</tr>
<tr>
<td>Rare</td>
<td>95 - 70</td>
<td>70 16</td>
<td>64 14</td>
</tr>
<tr>
<td></td>
<td>70 - 95</td>
<td>68 11</td>
<td>63 15</td>
</tr>
<tr>
<td>Common</td>
<td>95 - 70</td>
<td>61 13</td>
<td>67 16</td>
</tr>
<tr>
<td></td>
<td>70 - 95</td>
<td>67 11</td>
<td>64 15</td>
</tr>
</tbody>
</table>

Table 5.3: Means (in bold) and standard deviations from the analysis of subjects' confidence ratings in Experiment 6.
There was one highly significant main effect. This was the effect of Before/After manipulation ($F(1, 66) = 71.13, p < .001$). The mean confidence rating before receipt of the second piece of evidence was 66 (S.D. = 11), whilst the mean rating after was 76 (S.D. = 15). None of the other main effects approached significance. The single significant two-way interaction was that between Order and Before/After ($F(1,66) = 6.97, p < .02$). This interaction is shown in Figure 5.1. Tukey HSD tests for unequal sample sizes revealed significant differences between the means involved in this interaction. These means and significant differences are given in Table 5.4. There were no other significant results.

However, the three way interaction between Rarity, Order, and Before/After ($F(1,66) = 1.40, p > .24$), is shown in Figure 5.2.

![Figure 5.1: The significant interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 6.](image)

<table>
<thead>
<tr>
<th>Order</th>
<th>Bef/Aft</th>
<th>Mean Confidence</th>
<th>Std. Dev.</th>
<th>Sign. Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% - 70%</td>
<td>before 1</td>
<td>166 (11)</td>
<td>.0008</td>
<td>.0002</td>
</tr>
<tr>
<td></td>
<td>after 2</td>
<td>.0008</td>
<td>273 (14)</td>
<td>.0003 .01</td>
</tr>
<tr>
<td>70% - 95%</td>
<td>before 3</td>
<td>.0003</td>
<td>365 (10)</td>
<td>.0002</td>
</tr>
<tr>
<td></td>
<td>after 4</td>
<td>.0002</td>
<td>479 (15)</td>
<td>.01</td>
</tr>
</tbody>
</table>

Table 5.4: Means (in bold), standard deviations (in parentheses) and significant differences between means for the interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 6.
As may be seen from Figure 5.2, although the interaction between Order, Rarity and Before/After was not significant, the trend towards recency is more pronounced amongst subjects who received information about common features than amongst those who received information about rare features.

5.2.3.2 Expectation measures

A 2x2x3 mixed design Anova was used to analyse subjects' expectations about the unseen evidence. The between subjects factors were Rarity and Order. The within subjects factor was problem content. A total of 13 subjects failed to respond to at least one of the questions measuring expectation, so for the purposes of this analysis they have been excluded. Of the remaining 59 subjects there were 15 in the rare features/95% - 70% condition, 16 in the rare features/70% - 95% condition, 15 in the common features/95% - 70% condition, and 13 in the common features/70% - 95% condition. The means and standard deviations from this analysis are given in Table 5.5, whilst the full Anova table is given in Appendix 3.
There were two significant main effects. The first was for Rarity, $F(1,55) = 4.24, p < .05$. Subjects receiving information about a rare feature expected 65% (S.D. = 15) of members of the hypothesised set to possess the second rare feature. The equivalent mean for subjects in the common features condition was 72% (S.D. = 11). The second significant main effect was for problem content, $F(2, 110) = 9.04, p < .0003$. The mean expectation for the house scenario was 74% (S.D. = 18), for the engineer scenario it was 59% (S.D. = 25), and for the car scenario 72% (S.D. = 21). The only other effect which approached significance was the interaction between problem content and information order $F(2, 110) = 2.89, p < .06$. This interaction is displayed in Figure 5.3. Tukey HSD tests for unequal sample sizes revealed significant differences between means. These means and any significant differences between these means are given in Table 5.5.

### Table 5.5: Means (in bold), and standard deviations from the analysis of subjects' expectations in Experiment 6.

<table>
<thead>
<tr>
<th>Rarity</th>
<th>Order</th>
<th>House</th>
<th>Engineer</th>
<th>Car</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare</td>
<td>95-70</td>
<td>80</td>
<td>56</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>70-95</td>
<td>63</td>
<td>58</td>
<td>63</td>
</tr>
<tr>
<td>Common</td>
<td>95-70</td>
<td>81</td>
<td>58</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>70-95</td>
<td>71</td>
<td>65</td>
<td>82</td>
</tr>
</tbody>
</table>

Figure 5.3: The interaction between Order and problem content from the analysis of subjects' expectations in Experiment 6.
### 5.2.4 Discussion

As Experiment 6 was exploratory in nature the results of this experiment cannot be discussed in terms of predictions which have been confirmed or disconfirmed. However, the results of this experiment do have some things in common with the results of previous experiments. For example, as with previous experiments a significant main effect was found for the Before/After manipulation. Subjects were significantly more confident that the X hypothesis was the case after receipt of the second piece of evidence. Unlike Experiment 5 however, this experiment has produced an overall recency effect. Although subjects in both Order conditions were significantly more confident after receipt of the second piece of information than they were before receipt of that information, the final confidence ratings of those subjects in the 70% - 95% condition (i.e. subjects who received the strong information second) were significantly greater than the final confidence ratings of those subjects in the 95% - 70% condition (i.e. subjects who received the strong information first).

However, when the three way interaction between Order, Before/After, and Rarity is examined (see Figure 5.2), it can clearly be seen that the significant two way interaction is due to the trend towards recency present in the ratings of those subjects who received information about common features. There is no such trend present in the results of those subjects who received information about rare features. Although non-significant, this interaction suggests that subjects derived the subjective value of the pieces of information

<table>
<thead>
<tr>
<th>Order</th>
<th>Content</th>
<th>1 81 (18)</th>
<th>2 57 (28)</th>
<th>3 73 (24)</th>
<th>4 67 (16)</th>
<th>5 61 (22)</th>
<th>6 72 (19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% - 70%</td>
<td>house</td>
<td>0.0005</td>
<td>0.005</td>
<td>0.05</td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>eng</td>
<td></td>
<td>0.005</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>car</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70% - 95%</td>
<td>house</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>eng</td>
<td></td>
<td></td>
<td></td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>car</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.6: Means (in bold), standard deviations (in parentheses) and significant differences between means involved in the interaction between Order and problem content from the analysis of subjects' expectations in Experiment 6.
which they received from different sources. It was suggested in the introduction to Experiment 6 that those subjects who received information about features with high perceived base rates would focus on those features when assigning a subjective value to the pieces of evidence which they received. Accordingly, as there was no difference in the perceived base rate of these rare features, there would be no difference in the subjective value assigned to the two pieces of evidence. Such a suggestion fits with the finding that, regardless of information order, the confidence ratings of these subjects increased upon receipt of the second piece of information.

On the other hand, it was suggested that those subjects who received information about two common features would focus on the likelihood information when assigning a subjective value to the evidence. Because the likelihoods given were different, it was suggested that there would also be a difference in the subjective value assigned to the pieces of evidence. Such a difference in the subjective values assigned to the pieces of evidence would account for the recency effect observed in the common features half of the experiment.

Although there are several accounts of belief revision which can, post hoc, account for the results, no existing account is capable of making predictions about subjects' locus of attention when assigning a value to a piece of evidence. It is interesting that, despite over forty years of experimental work on order effects, no account has been given emphasising the importance to the effect of information order of how the subjective value of the evidence is derived. It is clear from these results that an adequate model of the belief revision process must account for both the derivation of subjective values for pieces of evidence and the effect of these subjective values on subjects' beliefs about the world. As was argued in the introduction to this experiment, such an account must be pragmatic in nature. This experiment suggests that, in this case at least, when there are competing sources of information available to the organism, that information which is most relevant will be used.
The analysis of subjects' expectations confirmed the results of the pre-test. A significant difference was found in subjects' expectations about the unseen evidence due to base rate of the features about which subjects received information. Those subjects who received information about common features had significantly higher expectations about the second piece of evidence than did subjects who received information concerning two rare features. This is reflected in the results of the pre test (see Table 5.1) which show that the perceived association between the common features used in this experiment was higher than for the rare features. However, there are several a priori reasons for suspecting that subjects' expectations played no role in the results of this experiment. Firstly, although significant, the observed difference in expectations was relatively small when compared to the differences in expectations found in cases where those expectations were claimed to be the crucial mediating factor. Secondly, if subjects' expectations were responsible for the results of this experiment, a significant two way interaction between the base rates of the features and their likelihoods should have been observed. Such a result was not obtained. This can be attributed to two things, the absence of a significant main effect for likelihood and the fact that only a partial order manipulation was used in this experiment - only the order of the likelihood was rotated, the order in which subjects were given information about the features remained the same across conditions. The third reason for doubting that differences in expectations led to this result is the effect of problem content on subjects' expectations. Although this factor did have a significant effect on subjects' expectations about the unseen evidence (as the results of the pre test suggested might be the case) no corresponding effect of problem content on subjects' confidence ratings was observed. This leads to the conclusion that expectation did not underlie the recency effect found amongst the results of those subjects given information about common features.

The final reason for suspecting the non-influence of subjects' expectations about the second piece of evidence given information about the first piece of evidence is connected with the point made in the Method section concerning the possible effects of differences in expectations between the Common and Rare conditions. It was pointed out that the results of Experiment 5 suggest that, in certain cases, the extent to which expectations about unseen evidence are exceeded by the actual value of that evidence
determines the extent to which subjects revise their beliefs in the light of that evidence. In this experiment, expectations were greater in the Common condition than in the Rare condition. This should have lead to greater upwards belief revision amongst subjects in the Rare condition. This was not the case.

It was also pointed out in the Method section that if the absolute value of the difference between subjects' expectations about the new feature given possession of the old feature and their expectations about the new feature on its own was taken as a measure of the Association between the features, then there was greater Association between the Rare features than the Common features. It was pointed out that the relationship between such a measure of Association and belief revision was not clear. However, given the fact that this measure seemed to indicate a strong, positive, relationship between the rare features used in this experiment it might be argued that subjects in the Rare conditions would have stronger expectations about the unseen evidence than would subjects in the Common conditions. In this case a prediction opposite to the one made above would be made: greater belief revision amongst subjects in the Common features condition. Once again, no such result was observed.

Before moving to a description of the next experiment however, the partial nature of the order manipulation used in this experiment deserves some discussion. Strictly speaking, a true order effect was not observed. This is because the Order manipulation consisted only of rotating the likelihoods with which the features were said to be possessed by instances of category X. In all cases subjects received information about the evidential features in the same order. This was done because it was the likelihood information which was of prime interest in this study. As the base rates of the features used was kept approximately constant within the base rate conditions, the non-reversal of feature order was not expected to affect the results of the experiment. The fact that the Order manipulation had no significant effect on subjects' expectations confirms the success of the pre-test.
5.3 EXPERIMENT 7

5.3.1 Introduction

Experiment 7 will be a direct follow-on from the previous experiment and, as such, will require little introduction. The finding of a trend towards recency with common features but not with rare features in the previous experiment, was interpreted as being due to subjects using different sources of information in assigning subjective values to the pieces of evidence which they received. Subjects in the rare features half of the experiment were claimed to have focused on the evidential features because of their rarity, whereas subjects who received information concerning common attributes were claimed to have focused more on the likelihoods. The problem with this interpretation of Experiment 6 is its speculative nature. However, this interpretation, although speculative, does lead to some very specific predictions. Experiment 7 will test one such prediction. This prediction is that the confidence ratings of subjects on a task which is structurally similar to that used in Experiment 6, but about the content of which subjects can have no background knowledge, will be exactly the same as the confidence ratings of subjects who received information about common features in Experiment 6.

The simplest way to dispense with the problem of subjects' background knowledge about the base rates of the features used, is to employ an abstract version of the task. Such a distinction, between abstract and concrete versions of a task, is to be found in the experimental literature on the Wason selection task (see Chapter 1 for an extensive discussion). In that literature, manipulations of the content of the task are also used to control for the background knowledge which subjects bring to bear on the experimental situation.

An abstract version of the task used in Experiments 3 - 6 is very easy to construct. Imagine being told that there exist two mutually exclusive categories X and Y. You are also told that there further exists an object O which belongs to one of these categories, and that it is your task to decide, in the light of evidence which will be provided, which of the
categories O is most likely to belong to. Firstly you are provided with information that O possesses two features, a and b. Next you are told that 95% of instances of category X also possess feature a, and that 70% of instances of X possess feature b. Given this evidence you must rate your confidence in the hypotheses.

How do you, in the absence of any information in background knowledge about the categories or the features, integrate the evidence which you possess with the available hypotheses? Although a subject, faced with this kind of task, has no background information about the problem content, she does have knowledge about likelihood expressions. From this knowledge she may infer that the fact that 70% of instances of category X possess feature b is evidence in favour of that category. Likewise, knowing that 95% of instances of X possess feature a is even stronger evidence in favour of category X. As it was claimed in the discussion of Experiment 6 that subjects who had been given information about common features focused on the likelihood with which instances of category X were said to possess the features, rather than the features themselves, the confidence ratings of subjects given an abstract version of the same task should be almost identical to those of subjects given information about common features.

Although the example of an abstract task given above consists of exclusive categories, X and Y, and a target object O which possessed features a and b, such a level of abstraction is neither necessary, nor desirable, in order to control for subjects' background knowledge. Instead, the experiment to be reported here will employ materials analogous to those already used, but abstract in the sense that subjects can have no background knowledge about the base rates of the features used. However, the levels at which instances of the focal category were said to possess the evidential features used in Experiment 6, will also be used in this experiment. The experiment to be reported here will also revert to the order manipulation used in Experiments 3, 4, and 5. In other words, a full order manipulation will be employed where all subjects will receive the same information, but the order of that information will be reversed for half of the subjects.
Predictions about the results of this experiment may be made at several levels. Firstly it is predicted that a recency effect will be found. Those subjects who receive the evidence in the strong/weak order will be significantly less confident, upon receipt of both pieces of evidence, that the focal hypothesis is the case than will subjects who receive the evidence in the weak/strong order. As the experiment to be reported here is a replication of one half of Experiment 6, but with abstract materials, a second, qualitative, prediction may be made. The interaction between information order and when subjects express their confidence from this experiment, should resemble more closely that half of the three way interaction between Order, Before/After and Rarity from Experiment 6 involving common features, than the half involving rare features.

5.3.2 Method

Subjects: 36 subjects took part in this experiment. 12 of these were male and 24 were female. Subjects were recruited by placing notices around the psychology department at the University of Plymouth. All subjects were paid for their participation. The mean age of subjects was 25.8. The oldest subject was 45 and the youngest was 18.

Materials: as with previous experiments, subjects received a handout which comprised of an instruction sheet and three problems. The instructions were as follows:

Accompanying these instructions is a series of three decision problems which require you to imagine yourself to be a linguist from the 25th century. Each of the problems concerns parts of speech (nouns, pronouns etc.) from a language spoken on the planet Zog. Zogese is a symbolic language based on combinations of concepts which approximate to earthling concepts such as diagonal lines, dots, squares, triangles, circles etc.

In each problem you will be given a description of a symbol and will be asked to decide which of two parts of speech the symbol represents. You will be given pieces of information relevant to each decision, and in each of the three problems you will be asked to rate your confidence in the two
alternatives in the light of this information. More detailed instructions on what
to do will accompany each of the problems.

Please bear the following points in mind throughout the experiment:

(1) Read each problem carefully and think hard before responding.

(2) Although there are some surface similarities between the problems, they all concern different parts of speech, so you should think carefully about each problem.

(3) Don't refer back to previous problems when working on the later problems.

(4) Each problem consists of two pages which are stapled together. It is important that you receive the information that they contain in the intended order. Don't go on to a later page until you are finished the one you are currently working on.

If you have any questions at any time please raise your hand and the experimenter will answer them. Thank you for agreeing to participate in this experiment.

The structure and content of the problems was also exactly the same as that used in previous experiments. Thus, for each problem, subjects initially received information such as the following:

You are currently studying a Zogese symbol. The symbol definitely represents either an adjective or a passive verb, and your task is to decide which of these two parts of speech the symbol actually represents. The symbol itself is triangular and there is a diagonal line running through the triangle.

You have the following piece of information:
95% of adjectives in Zogese have a triangular element.

Now look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that the symbol you are studying represents a passive verb and the other to complete certainty that it represents an adjective. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

certain that the symbol represents a passive verb  certain that the symbol represents an adjective

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of passive verbs in Zogese would you expect to have a triangular element?

(2) What percentage of adjectives in Zogese would you expect to have a diagonal line running through the main part of the symbol?

(3) What percentage of passive verbs in Zogese would you expect to have a diagonal line running through the main part of the symbol?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.
As may be seen from the above problem materials, subjects were once again asked about their expectations for each possible piece of evidence. This was done in order to keep the conditions in this experiment as similar as possible to those in previous experiments. Once again, as a manipulation check, only those answers about the piece of evidence which subjects were to receive will be analysed here. The second page of each problem contained the following information:

*Here is a second piece of information about the problem on the previous page:*

70% of adjectives in Zogese have a diagonal line running through the main part of the symbol.

Once again, there followed a rating scale where subjects were asked to indicate their confidence in each of the hypotheses.

As with Experiments 3, 4, and 5, the order in which subjects received pieces of evidence was manipulated. Thus, subjects received pieces of evidence in a strong/weak or weak/strong order. Accordingly, for the example given above, half of the subjects were initially told that 95% of adjectives in Zogese have a triangular element, whilst the other half were initially told that 70% of adjectives in Zogese have a diagonal line running through the main part of the symbol. As stated above, three problem contents were used in this experiment. In addition to the example given above, subjects were asked to decide whether a symbol, possessing a circular element and followed by a series of three dots, was a proper noun or a preposition, and whether a symbol, possessing a rectangular element and a line over the main part of the symbol, was an active verb or a pronoun. The full text of all the problems used is given in Appendix I.

*Design:* this was a 2x2x3 mixed design, with the between subjects factor being the order in which information was presented to subjects (Order). As with the previous experiment, the Order manipulation involved half of the subjects receiving information that 95% of instances of the X category possessed one feature followed by information that 70% of
instances of that same category possessed the second feature. The remaining subjects received this same information in the reverse order. For ease of exposition, therefore, the levels of the Order manipulation will be referred to as 95%/70% and 70%/95%. The within subjects factors were when subjects rated their confidence in the hypotheses (Before/After), and problem content. As with Experiments 4 - 6, the second measure was subjects' expectations about the second piece of evidence. As this measure was taken before receipt of the second piece of evidence, the Before/After factor is not involved here. This measure therefore had a 2 x 3 design. The design of both parts of the experiment is summarised in Table 5.6

<table>
<thead>
<tr>
<th>Task</th>
<th>Independent Variables</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Ratings</td>
<td>1. ORDER</td>
<td>2 - 95%/70%; 70%/95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 - Confidence Before</td>
</tr>
<tr>
<td></td>
<td>2. BEFORE/AFTER</td>
<td>Confidence After</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 - see Appendix 1</td>
</tr>
<tr>
<td>Expectations</td>
<td>1 and 3</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7: The independent and dependent variables involved in Experiment 7.

Procedure: subjects were run in groups of between ten and sixteen. In each condition three subjects received the problems in each of the six possible orders. They worked on their problems individually.

5.3.3 Results

Once again, in this experiment, subjects were asked to express their confidence on a 100mm line. At one end of the line was labelled "certain...X" and at the other end "certain...Y". For the purposes of the analysis which follows subjects' markings on the line were converted to scores on a 100 point scale ranging from 1 to 100. Each point on this scale corresponds to one millimetre on the line and the higher is a subject's score on this scale the more confident she is that the X hypothesis is the case.

One subject, in the 95%/70% condition, failed to complete at least one of the confidence measures. For the purposes of this analysis they have been excluded.
A 2x2x3 mixed design Anova was used to analyse subjects' confidence ratings. The between subjects factor was Order, whilst the within subjects factors were Before/After and problem content. The means and standard deviations from this analysis are given in Table 5.7, whilst a full Anova table is given in Appendix 3.

<table>
<thead>
<tr>
<th>Order</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjective/</td>
<td>Proper Noun/</td>
</tr>
<tr>
<td>95 - 70</td>
<td>Passive Verb</td>
<td>Preposition</td>
</tr>
<tr>
<td>67</td>
<td>16</td>
<td>66</td>
</tr>
<tr>
<td>70 - 95</td>
<td>62</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 5.8: Means (in bold) and standard deviations from the analysis of subjects' confidence ratings in Experiment 7.

The main effect of the Before/After manipulation was highly significant \( (F(1, 33) = 86.87, p < .001) \). Subjects' mean confidence before receipt of the second piece of information was 64 (S.D. = 14), whilst the mean rating after was 78 (S.D. = 15). Neither the main effect of information order \( (F(1, 33) < 1) \), nor the main effect of problem content \( (F(1, 66) < 1) \), were significant.

The only other effect which was significant was the two-way interaction between the Order and Before/After manipulations \( (F(1, 33) = 15.16, p < .001) \). The means and standard deviations involved in this interaction, as well as significant differences between means (tested for with Tukey HSDs for unequal sample sizes), are presented in Table 5.8. As may be seen from Figure 5.4 a significant recency effect was found in the confidence ratings of subjects in this experiment. As strong predictions were made at the outset about this interaction in relation to the three way interaction between Order, Before/After, and Rarity from Experiment 6, Figure 5.4 depicts the relevant means from both experiments.
Table 5.9: Means (in bold), standard deviations (in parentheses), and significant differences between means for the interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 7.

<table>
<thead>
<tr>
<th>Info Order</th>
<th>Bef/Aft</th>
<th>1 65 (18)</th>
<th>2 74 (18)</th>
<th>3 62 (10)</th>
<th>4 82 (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%/70%</td>
<td>before 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>after 2</td>
<td>.005</td>
<td></td>
<td>.0005</td>
<td></td>
</tr>
<tr>
<td>70%/95%</td>
<td>before 3</td>
<td></td>
<td></td>
<td>.0005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>after 4</td>
<td>.0005</td>
<td></td>
<td>.0005</td>
<td></td>
</tr>
</tbody>
</table>

| Figure 5.4: A comparison of the interaction between Order and Before/After from Experiment 7 (on the right) with the three-way interaction between Order, Before/After, and Rarity from Experiment 6 (on the left). |

Subjects’ expectations about the second piece of evidence were analysed using a 2x3 mixed design Anova. The between subjects factor was Order, whilst the within subjects factor was problem content. A total of four subjects from each of the between subjects condition failed to express their expectations for at least one of the problems. For the purposes of this analysis they have been excluded. The means and standard deviations from this analysis are given in Table 5.9, whilst a full Anova table is given in Appendix 3. None of the effects were significant. Of most interest was the main effect of information order (F(1, 26) < .2). Subjects in the 95%/70% condition expected 48% (S.D. = 18) of instances of the focal category to possess the other feature, whilst subjects in the 70%/95% condition expected 46% (S.D. = 12) of instances of the focal category to possess the other feature.
5.3.3 Discussion

A discussion of the results from this experiment is a relatively simple affair. As in previous studies, a significant main effect for the Before/After manipulation has been found. Once again, this means that, overall, subjects were more confident that the focal hypothesis was the case after receipt of the second piece of evidence. Also as predicted, was the significant interaction between Order and when Before/After. Although subjects in both order conditions were significantly more confident that the focal hypothesis was the case after receipt of the second piece of evidence, the final confidence ratings of subjects in the 70%/95% condition were significantly higher than those of subjects in the 95%/70% condition. An examination of Figure 5.4 reveals how similar the results of this interaction in the present experiment are to the common features half of the three-way interaction between Order, Before/After, and Rarity from Experiment 6.

One of the predictions made at the outset of this experiment was that there would be no significant differences in subjects' expectations about the unseen evidence (see pg. 227). This prediction has been confirmed. Confirmation of both predictions made at the outset (pg. 227) suggests that the account given of the results of Experiment 6 was indeed correct. Because the task used in this experiment was unfamiliar to subjects, the only source of information which they had available to them when relating evidence to hypotheses was knowledge about likelihoods. Most surprising is the fact that subjects were willing to rate one hypothesis as being more likely than the other in the light of such information. As Figure 5.4 demonstrates however, subjects' performance on this task was virtually identical to the performance of those subjects who received information about common features in Experiment 6. This suggests that those subjects in Experiment 6 did
not take the base rates of the evidential features, about which they received evidence, into account when relating evidence to hypothesis. The difference between the results of subjects who received information about rare features in Experiment 6, and those of the remaining subjects in that experiment, and in this one (also to be seen in Figure 5.4) suggests that the former were assigning subjective values to the evidence which they received based on a different source of information. As was pointed out in the discussion of Experiment 6, that source of information is likely to have been background knowledge about the base rates of the evidential features.

The results of this experiment, because of their similarity to those from subjects in the common features conditions of Experiment 6, suggest that subjects' expectations about the unseen evidence were not a factor in the judgements of those subjects. The similarity of these results to the previous ones also suggests that the partial order manipulation used in Experiment 6 was not a factor in the results. As Experiment 4 has demonstrated, reversing the order of the features leads to a recency effect only when there is a high degree of perceived association between the features. In other words, manipulating the order of the features has an effect only when knowledge about the relationship between the features is used as a source of information in integrating evidence and hypothesis. The results of this experiment, and the previous one, suggest that subjects do have access to other sources of information. The use of these other sources may also lead to order effects in belief revision. The problem for existing accounts is to specify what information is used, and when.

5.4 EXPERIMENT 8

5.4.1 Introduction

This section of Chapter 5 will contain the details of the final experiment to be described in this thesis. One of the aims of this experiment will be to demonstrate the differential use of information from background memory in assigning subjective values to pieces of evidence within the same task. In attempting such a demonstration, subjects will
be asked to form, and revise, beliefs in a pair of exclusive hypothesis in the light of diagnostic information relating to those hypotheses. This signals both a change in the experimental task, and a widening of focus.

In Experiment 1 subjects were asked to choose a piece of information in order to decide between two hypotheses. The order effect established in Experiment 3, and further explored in the subsequent experiments, was based upon the results of those subjects who chose pseudodiagnostic evidence to help them decide between the hypotheses. The question of how subjects use diagnostic information to decide between hypotheses has received little attention in this thesis. Before discussing the questions which a consideration of diagnostic evidence raises, a reminder of the differences between diagnostic and non-diagnostic versions of the task will prove useful.

Thus far, the majority of subjects in this thesis have been given a scenario containing two alternatives followed by some information relevant to one of those alternatives, and have then been asked to use that information to indicate their relative confidence in the alternatives. A typical example might be the subject told that her sister has just bought a new car which she knows to be a model X or a model Y but can't remember which. She does remember that the car possesses a radio and four doors. Next, the subject is told that 95% of car Xs possess a radio and 70% possess four doors. Her task is to decide how likely each of the alternatives is in the light of such evidence. From the point of view of traditional normative analyses of the task (Mynatt, Doherty and Dragan, 1993), the evidence which this subject has received is non-diagnostic - no decision can be made between the alternatives based on such evidence. This is because all of the evidence which the subject has received pertains to just one of the alternatives and therefore, the subject possesses no information upon which to base a comparison of those alternatives. However, as was pointed out in the discussion of Experiment 2, there exists the possibility of alternative normative analyses of the pseudodiagnosticity task which would assign informational utility to the pseudodiagnostic piece of information normally chosen by subjects on the task, and which the subjects in Experiments 3 - 7 of this thesis have been asked to use when deciding between the alternatives.
If, on the other hand, a subject given the same scenario, was told that 95% of model Xs possessed a radio and 25% of model Ys possessed a radio, then the subject could be said to possess diagnostic information (once again, from the traditional normative view). In other words, by the application of Bayes theorem (see Chapter I for a fuller discussion of this topic), the subject could work out how likely each of the alternatives is in the light of the evidence which she possesses. Although there exists considerable evidence (Doherty, Mynatt, Tweney and Schiavo, 1979; Mynatt, Doherty and Dragan, 1993) that subjects are insensitive to the diagnosticity of evidence when asked to choose information to help them decide between hypotheses (but see Chapters 1 and 2 for an alternative normative analysis of the pseudodiagnosticity task), there also exists evidence that subjects can make normatively correct use of such evidence once they possess it. For example, Beyth-Marom and Fischhoff (1983), based on a series of studies investigating subjects' use of diagnostic information, concluded that subjects are much better at using such information than they are at seeking it out. Likewise, Ofir (1988) found that subjects will use the likelihood ratio when they perceive it to discriminate between competing hypotheses. In a similar vein, Evans (1989) has pointed out how subjects are passively Popperian on Wason's (1960) 2 4 6 task. By this he means that although subjects do not seek out disconfirming evidence on the task, once they have received that evidence they are capable of putting it to use. Based upon such a consideration of the evidence, it is predicted that if subjects are given diagnostic information in a task such as the one used in earlier experiments, they will use it to provide confidence ratings which conform closely to those prescribed by Bayes theorem.

Nevertheless, the question of whether subjects make the normatively correct use of information is completely separate from the question of what processes are involved in the use of such information. In so far as subjects have been given non-diagnostic information in Experiments 3 - 7, this thesis could be said to be focused on the latter, rather than the former question. Although there will be a normative yardstick against which to measure subjects' performance in this experiment, the question still remains as to the processes underlying subjects' eventual confidence ratings. Before proceeding it should be pointed

237
out that the results of previous studies do suggest a normative account of the use of "pseudodiagnostic" information - its use should be related to the subject's prior expectations about that information.

The previous experiments in this thesis have suggested that the process of evidence interpretation is central to how subjects update their beliefs about a hypothesis in the light of new information. Experiments 6 and 7, in particular, demonstrated the use of different sources of information in background knowledge in assigning subjective values to pieces of evidence. These different sources of information were background knowledge about the base rates of the evidential features used and knowledge about the meaning of likelihood expressions. If, as predicted, subjects do use the diagnostic information which they receive to produce normatively correct confidence ratings, then it may be inferred that they are using background information about likelihood expressions in order to do so. In any probabilistic reasoning task, it is the level at which instances of the hypothesised categories possess a feature which should be taken into account when deciding which of the hypotheses is more likely, rather than the feature itself. This makes sense when one remembers that the hypotheses are being compared on a single feature. Thus it is the likelihood of the feature's occurrence which is important rather than the base rate of the feature itself.

However, both Experiments 2 and 6 have provided clear evidence that subjects are sensitive to, and use, information about the base rates of the features which they are reasoning about. How might such a sensitivity be expected to manifest itself on a task such as the one used in this thesis? Consider two groups of subjects - one group is given information concerning the likelihood that each of two hypothesised categories possesses a feature with a low base rate, whilst the other is given equivalent information concerning a feature with a high base rate. If the evidence is given sequentially, those subjects given information about a rare feature would be expected to use background information about base rates to assign a subjective value to the initial piece of evidence. However, upon receipt of the second piece of evidence, those same subjects would be expected to use the likelihood ratio as a basis for their confidence ratings. On the other hand, those subjects
who received information about a common feature would be expected to use the likelihoods in assigning subjective values to both pieces of evidence.

What is interesting about this hypothetical account is that it suggests that subjects who receive evidence concerning a rare feature will re-interpret the early piece of evidence in the light of the later piece of evidence. The first subjective value which they assign to the initial piece of evidence will be based on their background knowledge about base rates whereas the second subjective value assigned will be based on their knowledge of likelihood expressions. Experiments 3, 4, and 5 demonstrated that the subjective value of later evidence is dependent on inferences based upon both the initial piece of evidence and background knowledge. It is suggested here that, in certain cases, the meaning of early occurring evidence may be dependent upon inferences based on background knowledge and the later piece of evidence. Such a finding would be in stark contrast to current conceptions of information integration. For example, Massaro and Friedman (1990) in a paper reviewing and comparing current models of integration given multiple sources of information, claimed that the general stage model displayed in Figure 5.5 could be used to characterise all current models.

![EVALUATION](image1) → ![INTEGRATION](image2) → ![DECISION](image3)

**Figure 5.5**: Massaro and Friedman's (1990) general stage model of information integration.

Such a model is essentially forward working, and assumes that once a piece of evidence has been evaluated by the organism it is fed into the integration stage of the process. The product of the integration process is then fed into the decision stage. An example of such a model at work is that of Hogarth and Einhorn where each piece of evidence is integrated with previous evidence and the product of the integration is, itself, used in determining the effect of any subsequent evidence. This position is in stark contrast to the predictions made here, where it is expected that early evidence will be re-evaluated based on later evidence.
One way of testing these predictions is with an information order manipulation. If two groups of subjects are given the likelihood ratio for a feature, one group receiving information about a rare feature, and the other receiving information about a common feature, the initial, rather than final, confidence ratings of subjects should be dependent on information order. This is because subjects who receive an initial piece of information about a rare feature should be significantly more confident that the hypothesis to which it relates is the case than subjects who receive an initial piece of information about a common feature. However, if subjects use the likelihood ratio to make their final confidence ratings, then there should be only small differences in those final ratings due to the base rate of the features or information order.

5.4.2 Method

Subjects: 144 subjects participated in this experiment. They were recruited by placing notices around the University of Plymouth. 68 of the students were male and 75 were female. Their mean age was 23.2 years. The oldest subject was 45 and the youngest was 18. All subjects were paid for participation in the experiment.

Materials: each subject received a handout comprised of an instruction sheet and three problems. The instructions were as follows:

Accompanying these instructions is a series of three decision problems which require you, in the light of certain information, to rate your confidence in two alternatives and to answer some simple questions. Detailed instructions on what to do are contained in the problems but it would be very helpful if you could bear the following general points in mind throughout the experiment.

1) Read each problem carefully and think hard before responding.
(2) Although there are some surface similarities between the problems, they all concern different scenarios, so you should think carefully about each scenario.

(3) Don't refer back to previous problems when working on the later problems.

(4) Each problem consists of two pages which are stapled together. It is important that you receive the information they contain in the intended order. Don't go on to a later page until you have finished the one you are currently working on.

If you have any questions at any time, please raise your hand and the experimenter will answer them.

The first page of each problem was similar to the following:

Your friend has just bought a new house. You can't remember whether it's on street X or street Y. You do remember that the house has a swimming pool and a garage.

You have the following piece of information:

95% of houses on street X have a swimming pool.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend now lives on street X and the other to complete certainty that your friend lives on street Y. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the
greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Certain that your friend lives on Street X

 Certain that your friend lives on Street Y

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of the houses on street Y would you expect to have a swimming pool?

(2) What percentage of the houses on street X would you expect to have a garage?

(3) What percentage of the houses on street Y would you expect to have a garage?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

The second page of each problem contained the following information:

Here is a second piece of information about the problem on the previous page:

25% of houses on street Y have a swimming pool.

This was followed by a second rating scale.
All problems were of the same format as that given above. The first manipulation was that of information order. This was achieved by simply reversing the order in which subjects received the information about the hypotheses. One of the pieces of information always related to the focal hypothesis (category X) whilst the other always related to the non-focal hypothesis (category Y). It should be pointed out that the X hypothesis is referred to as the focal hypothesis because, in all cases, it is the hypothesis favoured by the likelihood ratio which subjects received. It is undoubtedly the case, however, that those subjects initially given a piece of evidence concerning the Y category will focus on the Y hypothesis until receipt of the second piece of information. For the above example, subjects given the focal evidence first received the information as it is presented here. Because the nature of the evidence given to all subjects meant that the focal hypothesis was always the one supported, the order manipulation in this experiment may be thought of as being analogous to the order manipulations in previous experiments. Accordingly, the focal evidence always offered strong support for the focal hypothesis whilst the non-focal evidence never offered as much support for the non-focal hypothesis and may, therefore, be thought of as offering weak evidence in favour of the focal hypothesis.

The second manipulation involved the rarity of the feature for which subjects were given the likelihood ratio. In the above example subjects were given information only about the rare feature although they were told that the target object possessed both a rare and a common feature. Half of the subjects were given information about the rare feature only and the other half about the common feature only. The features used were the same as those used in Experiment 5. The full text of the problems used is to be found in Appendix 1. The three scenarios used were the one presented above, one concerning an engineer-friend who works for one of two companies, and a problem concerning a friend who has bought either a model X or model Y car.

The third manipulation involved changing the likelihood ratio which subjects were given. All subjects were told that 95% of instances of the focal category possessed the relevant feature (this was always the strongest piece of information which subjects received). Half the subjects were also told that 25% of the non-focal category possessed
this feature whilst the other half were told that 60% of instances of the non-focal category possessed the feature.

**Procedure:** subjects were run in groups of between 4 and 10 in sessions without a time-limit. All subjects were paid for their participation in the experiment.

<table>
<thead>
<tr>
<th>Task</th>
<th>Independent Variables</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Ratings</td>
<td>1. ORDER</td>
<td>2 - FOCAL 1st;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FOCAL 2nd</td>
</tr>
<tr>
<td></td>
<td>2. RARITY</td>
<td>2 - COMMON;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RARE</td>
</tr>
<tr>
<td></td>
<td>3. LIKELIHOOD RATIO</td>
<td>2 - 95%/25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95%/60%</td>
</tr>
<tr>
<td></td>
<td>4. BEFORE/AFTER</td>
<td>2 - Confidence Before</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confidence After</td>
</tr>
<tr>
<td></td>
<td>5. PROBLEM CONTENT</td>
<td>3 different problem contents</td>
</tr>
</tbody>
</table>

Table 5.11: The independent and dependent variables involved in Experiment 8.

**Design:** this was a 2x2x2x2x3 mixed design. The between-subject variables were Order (Focal 1st vs. Focal 2nd), Rarity of the feature about which evidence was received (Rare vs. Common), and Likelihood Ratio (95/25 vs. 95/60). The within-subject variables were when subjects expressed their confidence (Before/After) and problem content. The design of the experiment is summarised in Table 5.10

### 5.4.3 Results

It will be remembered that subjects were asked to express their confidence on a 100mm line. At one end of the line was labelled "certain...X" and at the other end "certain...Y". For the purposes of the analysis which follows subjects' markings on the line were converted to scores on a 100 point scale ranging from 1 to 100. Each point on this scale corresponds to one millimetre on the line and the higher is a subject's score on this scale the more confident she is that the X hypothesis is the case.
Of the 144 subjects who participated, one subject, who received information about a common feature and a likelihood ratio of 95/60, failed to provide confidence ratings. Accordingly this subject will not be included in the analysis of subjects' confidence ratings which follow. Subjects' confidence ratings were analysed using a 2x2x2x2x3 mixed design analysis of variance. The between-subject factors were Order, Rarity and Likelihood Ratio. The within-subjects factors were Before/After and problem content. The means and standard deviations from this analysis are given in Table 5.11, whilst a full Anova table is to be found in Appendix 3.

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Rarity</th>
<th>Ratio</th>
<th>Order</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>House</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare</td>
<td>95/60</td>
<td>Focal 1st</td>
<td>66 16</td>
<td>68 14</td>
<td>64 16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Focal 2nd</td>
<td>42 12</td>
<td>42 10</td>
<td>75 13</td>
</tr>
<tr>
<td>95/25</td>
<td>Focal 1st</td>
<td>71 17</td>
<td>65 18</td>
<td>81 13</td>
<td>77 15</td>
</tr>
<tr>
<td></td>
<td>Focal 2nd</td>
<td>47 18</td>
<td>50 15</td>
<td>77 17</td>
<td>77 14</td>
</tr>
<tr>
<td>Common</td>
<td>95/60</td>
<td>Focal 1st</td>
<td>62 16</td>
<td>63 12</td>
<td>68 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Focal 2nd</td>
<td>46 10</td>
<td>40 11</td>
<td>70 12</td>
</tr>
<tr>
<td>95/25</td>
<td>Focal 1st</td>
<td>61 12</td>
<td>58 14</td>
<td>72 12</td>
<td>71 13</td>
</tr>
<tr>
<td></td>
<td>Focal 2nd</td>
<td>54 19</td>
<td>59 18</td>
<td>82 15</td>
<td>73 21</td>
</tr>
</tbody>
</table>

Table 5.12: Means (in bold) and standard deviations from the analysis of subjects' confidence ratings in Experiment 8.

The main effect of Order was highly significant (F(1, 135) = 16.92, p < .001). Subjects who received initial information about the focal hypothesis (the strong information) gave a mean rating of 68 (S.D. = 10) whereas subjects who received this information second gave a mean rating of 61 (S.D. = 9). The main effect of Likelihood Ratio was also significant, (F(1, 135) = 12.66, p < .001), as was the main effect of Rarity (F(1,135) = 3.98, p < .05).Subjects given a likelihood ratio of 95/25 had an overall rating of 67 (S.D. = 11) versus a rating of 62 (S.D. = 8) for those subjects given a likelihood ratio of 95/60. Subjects given information about rare features displayed a mean confidence rating of 66 (S.D. = 10) whereas subjects given information about the common features gave a mean rating of 63 (S.D. = 10).
The first of the within-subjects factors - Before/After - was also highly significant (F(1, 135) = 181, p < .001). The mean rating before receipt of the second piece of evidence was 56 (S.D. = 15), whilst the mean rating after was 73 (S.D. = 12). The effect of content was also significant (F(2, 270) = 3.64, p < .05). The mean rating for the house problem was 65 (S.D. = 12), for the engineer problem 63 (S.D. = 12), and for the car problem 66 (S.D. = 11). As an inspection of the data from the pre-test used to select features demonstrates (Chapter 4, Table 4.5), it was impossible to produce features with exactly the same frequency levels. Accordingly, slight differences due to content are to be expected. It should be pointed out, however, that although both the main effects of Rarity and problem content are significant, the effect is very small in each case.

Of the two-way interactions involved in the experiment, one - the interaction between Order and Rarity - approached significance (F(1, 135) = 3.33, p < .08) and one was highly significant - the interaction between Order and Before/After (F(1, 135) = 74.87, p < .001). These interactions are shown in Figures 5.6 and 5.7 respectively, whilst the means involved in these interactions are given in Table 5.12. Subsequent analyses of the simple effects in the interaction between Order and Before/After revealed a significant difference between the Order conditions when subjects expressed their confidence before receipt of the second piece of evidence (F(1, 135) = 35.43, p < .01) but not after (F(1, 135) = 1.63). Thus there was a significant difference in confidence expressed before the second piece of evidence, but not after receipt of that evidence. There was also a significant difference between before and after

<table>
<thead>
<tr>
<th>order</th>
<th>feature frequency</th>
<th>when confidence was expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rare</td>
<td>common</td>
</tr>
<tr>
<td>strong 1st</td>
<td>70 (11)</td>
<td>65 (9)</td>
</tr>
<tr>
<td>strong 2nd</td>
<td>62 (8)</td>
<td>61 (10)</td>
</tr>
</tbody>
</table>

Table 5.13: Means, and standard deviations (in parentheses), involved in the two-way interaction between Order and Rarity and in the two-way interaction between Order and Before/After. Both from the analysis of subjects' confidence ratings in Experiment 8.
measures when subjects received the Focal (X) information second ($F(1, 135) = 80.96, P < .01$) but not when they received that information first ($F(1,135) = 3.01, p > .05$). Subjects who received the Focal (X) information second were significantly more confident after receipt of the second piece of information than they had been before receipt of that information. The same was not true for subjects who received the Focal piece of information first.

**FIGURE 5.6:** The interaction between Order and Rarity from the analysis of subjects' confidence ratings in Experiment 8.

**FIGURE 5.7:** The interaction between Order and Before/After from the analysis of subjects' confidence ratings in Experiment 8.
Although the two way interaction between Likelihood Ratio and Before/After (displayed in Figure 5.8) was not significant (F(1,135) < 1), strong predictions had been made at the outset about differences between the means involved in this interaction. Tukey HSDs for unequal sample sizes were used to test for these differences. Their results are shown in Table 5.13.

<table>
<thead>
<tr>
<th>Likelihood Ratio</th>
<th>Before/After</th>
<th>1 54 (16)</th>
<th>2 70 (11)</th>
<th>3 58 (15)</th>
<th>4 77 (13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95/60</td>
<td>before 1</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>after 2</td>
<td></td>
<td></td>
<td>.001</td>
<td>.005</td>
</tr>
<tr>
<td>95/25</td>
<td>before 3</td>
<td></td>
<td>.001</td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>after 4</td>
<td>.001</td>
<td>.005</td>
<td>.001</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.14: Means (in bold), standard deviations (in parentheses) and significant differences between means for the two way interaction between Likelihood Ratio and Before/After from the analysis of subjects' confidence ratings in Experiment 8.

There were three significant three-way interactions. The first of these was the interaction between Order, Rarity and Before/After (F(1, 35) = 4.19, p < .05). An analysis of simple interaction effects revealed the interaction between Order and Before/After to be significant at both the rare level of the Rarity variable (F(1, 135) = 19.22, p < .01), and at the common level (F(1, 135) = 6.81, p < .025). Post hoc Tukey HSD tests for unequal sample sizes revealed significant differences amongst the means involved in this...
interaction. These are reported in Table 5.14 below, whilst the interaction is displayed in Figure 5.9.

![Graph showing mean confidence levels for different conditions]

**FIGURE 5.9:** The interaction between Order, Rarity, and Before/After from the analysis of subjects' confidence ratings in Experiment 8.

<table>
<thead>
<tr>
<th>Order</th>
<th>Rarity</th>
<th>Before/After</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>9</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>X 1st</td>
<td>rare</td>
<td>bef 1</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 2</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>comm bef 3</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 4</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>X 2nd</td>
<td>rare</td>
<td>bef 5</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 6</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>comm bef 7</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 8</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
</tbody>
</table>

**Table 5.15:** Means (in bold), standard deviations (in parentheses), and significant differences between means for the three-way interaction between Order, Before/After, and Rarity from the analysis of subjects' confidence ratings in Experiment 8.

The second significant three-way interaction was between Order, Likelihood Ratio and Before/After ($F(1, 135) = 9.59, p < .005$). Once again, an analysis of simple interaction effects revealed that Order interacted significantly with Before/After when subjects were given a likelihood ratio of 95/25 ($F(1, 135) = 5.20, p < .05$), and when the
ratio was 95/60 ($F(1, 135) = 22.18, p < .01$). Tukey HSDs for unequal sample sizes also revealed significant differences amongst the means involved in this interaction. These are shown in Table 5.15 below, whilst the interaction is displayed in Figure 5.10.

**FIGURE 5.10:** The interaction between Order, Likelihood Ratio, and Before/After from the analysis of subjects' confidence ratings in Experiment 8.

<table>
<thead>
<tr>
<th>Order</th>
<th>L’hood</th>
<th>Before/After</th>
<th>1 65</th>
<th>2 66</th>
<th>3 63</th>
<th>4 75</th>
<th>5 42</th>
<th>6 74</th>
<th>7 52</th>
<th>8 78</th>
</tr>
</thead>
<tbody>
<tr>
<td>X 1st</td>
<td>95/60</td>
<td>bef 1</td>
<td>.1</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 2</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>95/25</td>
<td>bef 3</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 4</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>X 2nd</td>
<td>60/95</td>
<td>bef 5</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 6</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>25/95</td>
<td>bef 7</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aft 8</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
</tbody>
</table>

Table 5.16: Means (in bold), standard deviations (in parentheses), and significant differences between means for the three way interaction between Order, Before/After, and Likelihood Ratio from the analysis of subjects' confidence ratings in Experiment 8.

Demonstrating once again that there were differences between the perceived frequency of the features, the interaction between Order, Likelihood Ratio and problem content was also significant ($F(2, 270) = 3.31, p < .05$). A further three-way interaction
involving problem content, Order, and Rarity approached significance ($F(2, 270) = 2.70, p < .07$).

5.4.4 Discussion

There were two predictions made at the outset of this experiment. The first of these was that all subjects would use the likelihood ratio which they received to provide confidence ratings close to those prescribed by Bayes theorem. The second prediction was that all of the differences between the means of the various groups of subjects in this experiment would occur amongst confidence ratings collected before receipt of the second piece of evidence. This prediction is of course limited to differences due to feature rarity and information order, as the first prediction implies differences in final confidence ratings due to the likelihood ratio.

The relationship between the results of the experiment and these predictions will be discussed in turn with the first topic of discussion being the normative accuracy of subjects' predictions. Overall, subjects' mean confidence rating in the 95/60 likelihood condition was 62, whilst the equivalent mean for those subjects in the 95/25 condition was 67. Although this difference was significant, it tells us nothing about subjects' accuracy as it is based upon the before and after ratings of all subjects. The most revealing result as regards subjects' normative accuracy is the non-significant interaction between Likelihood Ratio and Before/After (see Figure 5.8). Subjects in the 95/60 likelihood condition expressed a mean confidence of 70 after receipt of the second piece of evidence, whilst the equivalent mean for subjects in the 95/25 condition was 77. The difference between these means was significant. Given a likelihood ratio of 95/60, the probability, according to Bayes theorem, that the focal hypothesis is the case is .61, whilst with a likelihood ratio of 95/25, the probability is .79. Although subjects given a likelihood ratio of 95/60 seem to be over-confident that the focal hypothesis was the case, subjects given a likelihood ratio of 95/25 have used that information to produce confidence ratings almost exactly as prescribed by Bayes theorem.
The results of the three way interaction between Order, Likelihood Ratio, and Before/After (see Figure 5.10) demonstrate that information order did have some effect on subjects' final confidence ratings. With both likelihood ratios, those subjects who received initial information about the focal hypothesis displayed less final confidence that the focal hypothesis was the case than did subjects who received initial information about the non-focal hypothesis. This recency effect was not significant for either of the likelihood ratios. However, the trend towards recency was strongest amongst subjects who were given a likelihood ratio of 95/60. This was caused by the overconfidence of subjects in this condition who received the information about the non-focal hypothesis first. As predicted however, the majority of significant differences between the means involved in this interaction were between confidence ratings before receipt of the second piece of evidence.

Nevertheless, final confidence ratings can be said to have corresponded reasonably closely to what would be prescribed by Bayes theorem. In the context of, for example, research on utilisation of the base rate, the classic finding is that subjects grossly overestimate the posterior probability of the hypothesis through a neglect of the base rate when there is no clear causal link between the base rate and the probability of an outcome (although see Kohler, 1996, for a different reading of this literature). The difference between subjects' judgements and the normative answer is reduced considerably when that causal link is made clear (Tversky and Kahneman, 1980). Even with this improved performance however, the ratings of subjects in this experiment correspond more closely to the Bayesian norm than they did in the original Tversky and Kahneman study.

Ofir (1988) found that subjects used likelihood information to resolve inconsistencies present in the other information which they were given. When he presented subjects with a high base rate and hit rate (P(E/H), i.e. the likelihood of the data given the focal hypothesis) the pattern of judgements was similar to the normative curve. He concluded however, that this was due to subjects neglecting the false alarm (P(E/Ha)) and so subjects did not demonstrate an understanding of diagnosticity.
The probable reason for subjects relatively good performance on the task used in this experiment is its simplicity. Subjects were given only two pieces of information and the task was such that their relevance was apparent. The only way to resolve the inconsistency inherent in the categorisation task was to use both pieces of information. This of course, is not to claim that subjects, when cued to use both pieces of information, use them in a Bayesian way. The integration function is not Bayesian, as most of the studies in the area demonstrate (Pitz, Downing and Reinhold, 1967; Pitz, 1969; Roby, 1967; Shanteau, 1972; Peterson and DuCharme, 1967; Lopes, 1985; 1986). Thus, the prediction that final confidence ratings would approximate Bayesian solutions in this study was not motivated by a belief that the process is in any sense Bayesian, but rather by the observation that subjects can, in some cases, use information to derive ratings that are close to the norm.

The question of the process by which subjects came to use the information which they received in an apparently Bayesian manner is, as stated previously, separate from the question of their accuracy in using that information. It was predicted at the outset that there would be a differential use of sources of information available in background memory in assigning subjective values to the pieces of evidence. The result of interest in this regard is the significant interaction between Order, Rarity, and Before/After (see Figure 5.9). As predicted, there are large differences between confidence ratings involved in this interaction produced by subjects before receipt of the second piece of evidence. However, an examination of Table 5.14 where the means, and significant differences between means, involved in this interaction are displayed, shows that the only significant difference of interest is that between the initial confidence of subjects given information about the likelihood that a member of the focal (X) category would possess a rare feature, and the initial confidence of subjects informed of the likelihood that a member of the focal (X) category would possess a common feature. This finding leads to the inference that subjects who received the focal information first were using their knowledge about the base rates of the evidential features in assigning a subjective value to the initial pieces of evidence.
However, there is only one significant difference amongst the confidence ratings produced after receipt of the second piece of information. Subjects who received information about a common feature in the Focal 1st/non-Focal 2nd order were significantly less confident that the focal hypothesis was the case than were subjects who received information about a rare feature in the non-Focal 1st/Focal 2nd order. None of the other differences between the means involved in this half of the interaction were significant. It would seem that although subjects were using different sources of information before receipt of the second piece of information, all subjects were using the likelihood ratio after receipt of that piece of information.

That subjects were using different sources of information in assigning subjective values to the initial pieces of evidence which they received is further evidenced by a comparison of subjects' initial confidence ratings in the interaction between Order, Rarity and Before/After (see Figure 5.9) and the interaction between Order, Likelihood Ratio and Before/After (see Figure 5.10). As has already been pointed out, the former interaction suggests that subjects initially given evidence about the focal hypothesis were using information about the base rates of the evidential features in assigning a subjective value to the piece of evidence which they received. However, there is no such significant difference between the initial confidence ratings of those subjects who received information about the non-focal hypothesis first. In other words, these subjects do not seem to have been sensitive to the base rates of the evidential features about which they were given information. It is not the case that the initial confidence ratings of these subjects are completely heterogeneous, however. An examination of the three way interaction between Order, Likelihood Ratio and Before/After (see Figure 5.10) reveals a significant difference between the means of subjects given an initial piece of information about the non-focal hypothesis. This difference is due to the Likelihood Ratio manipulation. Thus, those subjects initially told that 25% of instances of the non-focal category (the Y category) possessed either the common or rare feature were significantly more confident that the focal hypothesis (the X hypothesis) was the case than were subjects initially told that 60% of instances of the non-focal category possessed either the common or rare feature. Accordingly, subjects who received evidence about the non-focal
hypothesis first were more sensitive to differences in the likelihood attached to that piece of evidence than they were to the base rate of the feature which it concerned.

This result suggests that subjects were using different sources of information to assign subjective values to the initial pieces of evidence which they received. However, all subjects seem to have predominantly used the likelihood ratio in making their final confidence rating. This is interesting because it suggests that subjects who were initially told that 95% of instances of the focal category possessed the rare feature were forced to re-evaluate that first piece of evidence in the light of the second piece of evidence which they received. This means that, in certain cases, even when a piece of information has been evaluated and a decision has been made about how that evidence relates to the hypotheses, that piece of information may have to be re-evaluated, and a new subjective value fed into the integration process. Such a re-evaluation is necessary in this experiment because once the subject possesses a full likelihood ratio than the rarity of the feature on which the hypotheses are being compared is of no informational value. So, although the rarity of the features was used by subjects in assigning a subjective value to the evidence when they possessed just one piece of evidence, such information was not relevant once the subject was in possession of the full likelihood ratio. As has been seen, subjects made very good use of the likelihood ratios which they were given.

Such an interpretation of these results is further backed up by a specific consideration of how they relate to the belief adjustment model of Hogarth and Einhorn. It has been repeatedly pointed out that their predictions, about order effects for the type of task used in the experiments contained in this thesis, rely on both the subjective values of the pieces of information which subjects receive and the difference in these subjective values. If the initial confidence ratings produced by subjects is taken as a rough index of the subjective value for the pieces of information which those ratings are based upon, it will be seen that the difference in subjective values between the pieces of information which subjects received is almost twice as large with rare features as it is with common features (see Figure 5.9). However, the recency effect produced in both halves of the three way interaction between Order, Rarity, and Before/After, is exactly the same. This result
can only be explained in terms of subjects who received an initial piece of evidence concerning a rare feature, re-evaluating that initial evidence.

Therefore, not only has this experiment demonstrated the differential use of sources of information from background knowledge within a belief revision task, it has also shown that the process of assigning a subjective value to a piece of information is reversible. This contrasts with an assumption which underlies all existing models of information integration and is nicely summed up by Massaro and Friedman (1990) when they claim that the scale value produced by the integration stage has no "memory" of how it was obtained. This experiment has demonstrated that initial evidence is held in memory and may be subjected to a second stage of evaluation, if later occurring evidence suggests that the initial evaluation was inadequate.

Although this experiment is far from enabling a complete account of how subjects integrate evidence with hypothesis, it does suggest that the process is more flexible than was previously suspected. It also has a bearing on what information is represented by the subject during the process, and how this representation is structured. These issues will be discussed in the next, and final chapter of the thesis. The discussion will be relatively conservative however. Whilst Experiment 8 suggests some very interesting avenues for further research, and asks some questions which current accounts of the belief revision process seem unable to answer, it should be seen as a starting point rather than an ending.

5.5 GENERAL DISCUSSION

The experiments described in this chapter have strengthened the case for a model of the belief revision process which includes a strong pragmatic component. These experiments have also cast further doubt on those existing models of the process which do not possess a pragmatic element. Although the subject of all three experiments is the question of how background knowledge is used in assigning subjective values to pieces of evidence, they differ slightly in their focus. Experiments 6 and 7 constitute an initial
demonstration of how the meaning, or strength, of any piece of evidence is determined by which of its components is focused on, whilst Experiment 8 suggests that subjects' focus may switch within the same task. With this in mind, it will be worthwhile, initially at least, to discuss Experiments 6 and 7 separately from Experiment 8.

The results of Experiments 6 and 7 force a modification of the sketch of the belief revision process offered at the end of Chapter 4. These experiments were based on an analysis of the information available to subjects engaging in the task used in Experiments 3, 4 and 5. Three sources of information were identified: information about the level of association between the evidential features; information about the base rates of the features; and information about the meaning of likelihoods. Experiment 6 was designed to investigate the differential use of the last two sources of information. Subjects were given information about either two rare, or two common, features. One possibility, recognised at the outset, was that subjects who received information about two common features would be more likely to produce a recency effect than would subjects who received information about two rare features. That is, with common features, the effect of an Order manipulation would be more likely to result in significantly higher final confidence ratings when the stronger piece of evidence was received second rather than first. No such effect of information order was predicted amongst the results of subjects who received information about rare features. This suggestion was based on the assumption that subjects given information about rare features would be more likely to use background information about the base rates of the features in assigning a subjective value to the evidence than would subjects who received information about two common features. As the features used were of equal rarity (and, therefore, likely to be equally diagnostic of the objects to be categorised), use of base rate information would minimise the contrast between the subjective values of the evidence. On the other hand, subjects who received information about two common features were expected to use the levels at which members of the focal category possessed the features in assigning subjective values to the evidence. As there was a difference in these levels, a difference in the subjective values assigned to the pieces of evidence was expected. It was suggested that this difference in subjective values would lead to a recency effect.
The results of Experiment 6 supported the predictions made at the outset. Although non-significant, the three way interaction between Order, Rarity and Before/After (see Figure 5.2) revealed a trend towards recency amongst those subjects who received evidence about common features but not amongst the results of subjects who received information about rare features. However, the results of Experiment 7 offer much stronger support for these predictions. Subjects in Experiment 7 behaved in the same manner, on an abstract version of the task, as had subjects in Experiment 6 upon receipt of evidence concerning two common features. Taken together, these experiments lend strong support to initial predictions concerning the differential use of information in assigning subjective values to pieces of evidence.

The confirmation of these predictions is interesting for two reasons. Firstly, they indicate, once again, the importance of pragmatics to the occurrence of an order effect. This experiment, along with those in Chapter 4, indicates that the process of assigning a subjective value to a piece of evidence is central to the effect of that evidence on the beliefs of the organism. Once again, existing accounts of the belief revision process are inadequate in this respect. All existing accounts (e.g. Anderson, 1981; Lopes, 1987; Hogarth and Einhorn, 1992) assume a subjective value for the pieces of evidence which subjects receive. They are incapable of accounting for the derivation of that subjective value. This may clearly be seen in Figure 3.1 from Chapter 3 which represents Shanteau and Nagy's (1984) conceptualisation of models of information integration. Although their conceptualisation allows for a separate stage for stimulus evaluation, it is clear from the research described in Chapter 3 that almost all experimental effort has gone into a description of the integration stage of the process. However, it is equally clear, from the results of Experiment 6, that what Shanteau and Nagy term the stimulus evaluation stage of the process, is at least as crucial to the final belief, as is the integration stage. Although, as was pointed out in the introduction to Experiment 6, there are several existing models which may account for the trend towards recency (significant in the case of Experiment 7, insignificant in the case of Experiment 6) observed in these experiments, these models are incapable of predicting such results. No existing model makes any predictions about how a
piece of evidence comes to have a meaning for the subject. Accordingly, in the absence of information about the subjective value of the evidence, they are unable to make \textit{a priori} predictions about when a recency effect will occur. Simply put, Hogarth and Einhorn, for example, predict that a difference between the subjective values of the evidence which subjects receive will lead to recency. However, because they offer no account of how these subjective values are derived, their model is incapable of predicting when there will be a difference in subjective values and, by extension, of predicting when recency will occur.

The second point of interest in these findings is their implications for any pragmatic component of belief revision. In the general discussion of the experiments contained in Chapter 4, it was claimed that the assignment of a subjective value to a piece of evidence was an inferential process where the selection of the premises upon which inferences were made was pragmatically determined. It is still claimed, based upon the results of the experiments contained in this chapter that the process is an inferential one. However, the range of sources of information available to the subject has been broadened by these experiments. Thus far, the use of three different sources of information has been experimentally demonstrated. The question becomes one of determining which source of information will be used in any given context.

As was discussed both in Chapter 1 and in the introduction to Experiment 6, Sperber and Wilson's (1986; 1995) Relevance Theory predicts the differential use of background information depending on context. In a similar vein, Shanon's (1988) critique of Tversky's (1977) model of similarity (discussed in Chapter 3) rests on the recognition of the importance of background knowledge to the selection of features upon which to compute similarity. How is a single source of information chosen from the many competing sources in background knowledge? The effect/effort trade-off described by Sperber, Cara and Girotto (1995) goes some way towards answering this question.

The effect/effort trade-off claims that only that information which is most relevant to the organism in any given context will become important. Relevance is determined by how much effect any piece of information has for the organism, balanced against how
much processing effort that information requires. Such a notion of cognitive economy may be used to account for the results of both Experiments 6 and 7 and the experiments in Chapter 4. In each experiment reported, there seems to have been one source of information which was of primary importance to subjects. In Experiments 3, 4, and 5 this was information in background knowledge about the level of association between the evidential features used. In Experiment 6, information about the base rates of the features used or knowledge about likelihood expressions, was important.

These three sources of information seem to be arranged in a hierarchy. In Experiments 3, 4, and 5, even though information about the base rates of the features used, and knowledge about likelihood expressions, was available to the subject, knowledge about associations between features was most important. Experiment 6 demonstrated that in the absence of strong associations between features, information about the base rates of the evidential features was most commonly used - but only if the base rates of the features were low. When the base rates of the features were high - as in the common features conditions - or was absent - as in Experiment 7 - information about the meaning of likelihood expressions was used in making the judgement.

Why the different sources of information should be so arranged is not completely clear. It is likely however, that a major factor underlying their arrangement is the fact that the background knowledge invoked becomes more specific to the context as one travels up the hierarchy. Knowledge about the relationship which exists between two features of an object constitutes more specific knowledge about the object than does knowledge about the base rates of its individual features, or indeed, than does no knowledge about its features. It is also clear that Sperber, Cara, and Girotto's (1995) effect/effort trade-off must be invoked in accounting for the fact that possessing background information about common features of an object results in the same pattern of results as possessing no background knowledge about the object. In the former case information about the likelihood that instances of the focal category possess the features, on its own, must be easier to represent and just as informative, as likelihood information and base rate information combined.
Experiment 8 demonstrates the differential use of two of the sources of information, discussed above, in the assignment of subjective values to pieces of information within the same task. Eight groups of subjects were given two pieces of information, comprising a likelihood ratio, concerning a pair of hypotheses. The order in which subjects received this information was manipulated, as were both the rarity of the feature to which the likelihood ratio related and the value of the likelihood ratio itself. After receipt of both pieces of information, subjects displayed a sensitivity to the actual value of the likelihood ratio and produced confidence ratings similar to those which would be prescribed by a consideration of Bayes Theorem. However, before receipt of the second piece of information, subjects' confidence ratings reflected both the rarity of the feature about which they had received evidence and the likelihood that instances of the category to which that evidence related possessed the feature. Interestingly, such differences were not apparent in subjects' final confidence ratings. It was argued that this result suggests that subjects were, in some cases, re-evaluating the earlier piece of evidence which they had received in the light of the later evidence. The result of most interest is the finding of a significant difference amongst the initial confidence ratings of those subjects told the percentage of instances of category X which possessed a rare feature and those told the percentage of instances of category X which possessed a common feature. The finding of such a difference suggests that subjects were using information about the rarity of the features in order to assign a subjective value to the pieces of evidence. However, upon receipt of the second piece of information (the percentage of instances of category Y which possessed either the rare or common feature) there was no such significant difference between the final confidence ratings of subjects in these groups. This suggests that those subjects who had assigned a subjective value to the first piece of evidence based upon the feature it concerned were forced to re-evaluate that evidence in the light of the later evidence. As there was just one significant difference between the final confidence ratings of the subjects in these groups it may be inferred that they made their final confidence judgement based primarily on the likelihood ratio rather than on the feature for which the likelihood ratio was given.
As stated previously, this finding contradicts the view of information integration (best expressed by Massaro and Friedman, 1990) normally held in the literature. This view holds that once a piece of evidence is evaluated and integrated with the hypotheses, it cannot be re-evaluated. In Massaro and Friedman's terms, a subjective value has no memory for the piece of evidence from which it was derived. If, as has been demonstrated, a piece of evidence may be re-evaluated, then some conclusions may be drawn about how subjects represent the task used in this thesis. At the very least, subjects must be holding in working memory a representation of the earlier evidence which they have received as well as a representation of the strength of that piece of evidence. This would facilitate the re-evaluation of that earlier piece of evidence in the light of later evidence. It is also likely that subjects are capable of focusing on different parts of their representation of the earlier piece of evidence. Thus, later evidence may be thought of as manipulating the locus of subjects' attention and, therefore, the meaning of earlier evidence.

5.6 SUMMARY

This chapter has described three experiments. The first of these suggested a differential use of information from background knowledge in assigning subjective values to pieces of evidence. Subjects received information about either common or rare features of the object to be classified. The weak prediction was made that information about common features would lead to a recency effect, whereas information about rare features would not. The findings suggested that these predictions were correct and were interpreted in terms of subjects in each condition using different sources of information in assigning a subjective value to the evidence. Those subjects who received information about rare features used that information about the rarity of the features in determining the value of the evidence. Accordingly, as both features were rare, there was no contrast in the strength of the pieces of information. However, those subjects who received information about common features of the objects used information about the meaning of likelihood expressions in assigning a subjective value to the evidence. Accordingly, the difference in the likelihoods that instances of the focal category possessed the evidential features, which had not been used by subjects in the rare features condition, resulted in a difference in the
subjective values of the pieces of evidence. This difference in subjective values led to the observed recency effect.

Such an interpretation of these results received support from the results of Experiment 7 which employed an abstract version of the task used in previous experiments. This experiment replicated Experiment 6 but used abstract materials and a full order manipulation (only a partial order manipulation was used in Experiment 6). The pattern of results for subjects in Experiment 7 was almost exactly the same as that of those subjects who had received information about common features in Experiment 6. This indicates that the interpretation of Experiment 6 was correct in claiming that subjects who received information about common features did not use information, stored in background knowledge, about the base rates of those features, to inform their confidence ratings of the hypotheses.

The results of these first two studies were interpreted as being damaging for existing accounts of the belief revision process. Any explanation of these findings requires a model which will explain how subjective values are assigned to pieces of evidence as well as how these subjective values are integrated, and such a model does not currently exist. This is because the results of Experiments 6 and 7 suggest that the process of assigning a subjective value to a piece of evidence, in many cases, determines the occurrence of an order effect. The results are also interesting in terms of explaining the influence of pragmatics on the belief revision process. They suggest that there are several sources of information available to the subject when assigning a subjective value to a piece of evidence. Along with the experiments contained in Chapter 4, they also suggest that these sources of information have differing degrees of importance to the subject. It was argued that the differential use of these sources of information is best explained by Sperber et al's trade-off between cognitive effect and cognitive effort.

Experiment 8, on the other hand, was an attempt to demonstrate the differential use of sources of information from background memory in assigning subjective values to pieces of evidence within a belief revision task. Subjects were given diagnostic evidence
concerning a pair of hypotheses in sequential order. There was a significant difference found due to changes in the likelihood ratio which subjects received. Subjects used these likelihoods in a manner approximating the use prescribed by Bayes theorem and, although there was a trend towards recency present in the data, the effect of information order did not significantly affect subjects' use of the likelihood ratio.

The differential use of sources of information was observed in this experiment. There were significant differences observed in subjects' initial confidence ratings which were ascribed to use of both background information about the base rate of the evidential features and background information about the meaning of likelihood expressions. However, there was only one significant difference observed between subjects' confidence ratings after receipt of the second piece of evidence. In general, subjects final confidence ratings were less varied than were their initial confidence ratings. This was explained as being due to the use of the likelihood ratio in comparing the hypotheses upon receipt of both pieces of evidence.

The results of this experiment were interpreted as providing evidence that subjects are capable of using the likelihood ratio as prescribed by Bayes theorem in very simple problems. It was not claimed that the integration function was Bayesian however. Secondly, this experiment was interpreted as providing evidence that the evaluation of a piece of evidence is reversible. Early occurring information may be re-evaluated in the light of later occurring information. Such a finding contradicts current views of belief revision and information integration.

The failure to find significant recency effects was also discussed. This was despite the fact that there were highly significant differences found between subjects' initial confidence ratings. Failure to find increased recency in those conditions where the difference in subjects' initial confidence ratings due to information order was twice as large as in the other half of the experiment was interpreted as further evidence against the notion that the size of a recency effect is determined by the degree of contrast between the pieces of evidence which subjects receive. Once again, it was argued that the effects
observed are explainable only in terms of the re-evaluation of early occurring evidence in the light of later evidence.
CHAPTER 6 - OVERVIEW

6.1 INTRODUCTION

This, the final chapter of the thesis, has four aims. The first of these is simply to summarise the results of the eight experiments described in Chapters 2, 4 and 5. These experiments fall naturally into three groups: experiments on information selection; experiments on expectation and belief revision; and experiments on the role played by selective attention in assigning subjective values to pieces of evidence. Accordingly, each of these groups of experiments will be summarised separately. The second aim of this chapter is to place the results of these experiments in context. Each summary section will, therefore, both outline the novelty of the experimental results which they summarise, and reflect upon how these results are to be integrated with the existing experimental literature. The third aim of this chapter will be to specify how the results of the thesis sit with current accounts of the role of background knowledge in both information selection and belief revision. The fourth, and final, aim of this chapter is to comment on the relationship between belief revision (normally construed as a subset of decision making) and hypothesis testing (which until recently was considered to be a problem for investigators of reasoning). It will be argued that recent empirical and theoretical work, and the results of the experiments contained in this thesis, suggest that many of the processes underlying these phenomena are the same.

6.2 A SUMMARY AND INITIAL DISCUSSION OF THE EXPERIMENTS ON INFORMATION SELECTION

The first two experiments in this thesis were concerned with subjects' use of background information concerning subjective probability when selecting evidence to help them decide between two hypotheses. These experiments were motivated by the review of the literature on information selection presented in Chapter 1. One of the conclusions of that review was that although there exist several theoretical accounts of information selection which suggest the importance of subjective probability to subjects' selection of
information for hypothesis evaluation (Evans and Over, 1996a&b, Oaksford and Chater, 1994) there is insufficient empirical work to discriminate between these various accounts. A second conclusion from the review presented in Chapter I was that the pseudodiagnosticity task (Doherty, Mynatt, Tweney and Schiavo, 1979; Mynatt, Doherty and Dragan, 1993) is a suitable instrument for testing intuitions about the role of subjective probability in hypothesis testing.

Using the pseudodiagnosticity paradigm, Experiment I attempted to establish a basic effect of subjective probability on subjects' patterns of information selection. The pseudodiagnosticity paradigm involves giving subjects a scenario where a target object possessing two features has to be categorised as belonging to one of two categories. The evidence which subjects receive consists of the likelihoods that instances of each of these categories possess the features possessed by the target object. As the specific version of the task which was used was taken from Mynatt et al, all subjects received an initial piece of information following receipt of the scenario. In order to examine the effect of subjective probability the rarity of the feature which this initial piece of information concerned was manipulated. A second aim of Experiment I was to examine the effect of the strength of this initial piece of information on subjects' subsequent information selections. Accordingly, one group of subjects were told that 65% of instances of one of the categories possessed one of the evidential features, another group that 95% of instances possessed the feature, whilst a third group were simply told that instances of one of the categories possessed the feature.

The results of Experiment I showed no effect for the rarity manipulation. The only effect of interest was the difference in information selection caused by manipulating the strength of the initial piece of information which subjects received. Subjects given information in propositional form (i.e. told that instances of one of the categories possessed the evidential feature) were significantly more likely to select further information about the category to which the initial piece of information related (further $H_x$ information) than were subjects who received information in the form of probabilities. It was argued that by giving subjects information in propositional form, the structure of the
task had been changed. Accordingly, subjects who received information in propositional form may have decided that further $H_x$ information was the most informative because of a consideration of all possible selections and their consequent outcomes. It was argued, however, that such an explanation for the results was unlikely and that subjects' tendency to select further $H_x$ information was more likely to be due to attentional factors. Specifically, it was argued that the initial piece of information which subjects received led them to focus on further information about the hypothesis which they were currently considering.

Experiment 2 constituted an attempt both to tease apart these possible explanations for the significant result found in Experiment 1, and to demonstrate an effect of subjective probability on the task. In Experiment 2 all subjects were given in evidence in propositional form. That is, all subjects were initially told that instances of one of the categories possessed one of the features said to be possessed by the target object. The remaining evidence, from amongst which subjects could choose was also expressed in terms of possession, or non-possession, of the evidential features. For half of the subjects the initial piece of evidence which they received concerned a rare feature whilst for the remaining subjects this initial piece of evidence concerned a common feature. Owing to the structure of the propositional version of the task, it was felt that one factor which may have contributed to the results of Experiment 1 was whether subjects made the pragmatic inference that there was a piece of evidence available to them which would be helpful in deciding between the alternatives. This was the case because the utility of each piece of evidence available was dependent on the actual value of that evidence. Because subjects did not have any information about these values, it is possible that some subjects may not have assumed that the evidence available was useful. In Experiment 2 half of the subjects in each of the Rarity conditions were told that there was, at least, one piece of evidence which, regardless of its outcome, would enable them to decide with certainty between the alternatives.

The results of Experiment 2 were, on the whole, as expected. Subjects who received an initial piece of evidence concerning a common feature of the object to be
categorised were significantly more likely to select further H\textsubscript{x} information than were subjects who received an initial piece of evidence about a rare feature. This result is a clear demonstration of the importance of subjective probability to subjects’ selection of information for hypothesis testing. The rarity of the feature about which subjects’ initially received some information determined the further information which they selected. In their patterns of evidence selection subjects demonstrated a sensitivity to the probabilistic structure of the task. Due to this probabilistic structure, initial information about a common feature meant that further H\textsubscript{x} information was likely to be more informative than information about the initial feature in relation to category Y. Conversely, given information about a rare feature, information about that rare feature in relation to category Y was likely to be more informative than further H\textsubscript{x} information.

Also as expected, there was no effect of whether subjects were told that at least one piece of information would allow them to decide between the alternatives with certainty. These results were interpreted as suggesting that subjects did, in any case, infer that the evidence available to them on the task was useful and, more importantly, that subjects considered the possible pieces of evidence in the order in which they were suggested by the problem content and settled on a piece of evidence which they felt would enable them to discriminate between the hypotheses.

A normative analysis of the experimental task used in Experiment 2 was also presented. Like the recent analyses of the selection task presented by Kirby (1994a), Oaksford and Chater (1994), and Evans and Over (1996a&b), this analysis relied on the concepts of subjective probability and subjective expected utility. It was demonstrated that subjects’ pattern of information selection on the task used in Experiment 2 closely corresponded to the prescriptions of the normative analysis. Thus when subjects received an initial piece of information concerning a common feature of the object to be categorised, the normatively correct selection was further information about the hypothesis to which the initial piece of information related. When the initial piece of information concerned a rare feature, however, the information to be gained from the selection of
information concerning whether instances of the alternative category possessed that feature was greatly increased.

The normative analysis presented in Chapter 2 was used, in conjunction with notions of relevance, to explain the pattern of results found in Experiment 2. This explanation may be thought of as constituting a model of subjects' behaviour on the task used in that experiment. Essentially the model is one of constrained rationality. By this it is meant that although the model assumes that subjects are sensitive to the probabilities and utilities of the various outcomes which are possible on the task, the extent to which they represent, or consider, these outcomes (and hence the utilities which subjects assign to them) is constrained by their limited cognitive resources. This constrained rationality model claims that subjects' patterns of information selections in Experiment 2 are to be captured by a consideration of subjects' knowledge about the probability of the task features, the extent to which they consider various possible outcomes, and the utilities which they assign to these outcomes. Crucially, the assignment of utilities is seen as being dependent on subjects' limited representational resources. Thus, the model claims that, in some cases, low utility (or no utility) will be assigned to particular outcomes because the subject has not considered them, either at all, or to any great extent.

This constrained rationality model is interesting for several reasons. Firstly, it is in line with several recent models of both similar (Klayman and Ha, 1987; Kirby, 1994a; Oaksford and Chater, 1994; Evans and Over, 1996a&b) and different (Anderson, 1990; Cheng and Novick, 1990; Corter and Gluck, 1992) cognitive processes. All of these models have in common the emphasis which they place on the organism's knowledge of subjective probability. Secondly, the constrained rationality model suggests that an alternative normative analysis of the pseudodiagnosticity paradigm is possible. The traditional analysis of the paradigm (Doherty, Schiavo, Mynatt and Tweney, 1979; Mynatt, Doherty and Dragan, 1993) would claim that failure to choose information relevant to the alternative hypothesis about the feature for which you already possess some information is a violation of the prescriptions of Bayes theorem. Interestingly, Stephen Stich (Stich, 1990) has claimed that results on the pseudodiagnosticity task demonstrate the irrationality
of human subjects. However, if the task is thought of as one about which subjects have expectations based on background knowledge about subjective probabilities, then Stich's claim is considerably weakened. Although this thesis has not demonstrated the use of knowledge about subjective probabilities on a standard version of the pseudodiagnosticity paradigm (Experiment 1, where such evidence might have been obtained, was underpowered), it has suggested that an alternative normative analysis is possible.

The third, and perhaps most interesting, point of interest in the constrained rationality model is its apparent psychological plausibility. In Chapter 1 Oaksford and Chater's model of the selection task was considered in considerable detail. This model was criticised on several grounds, but chiefly for its lack of psychological realism. For example, its claim that subjects engage in more processing on some versions of the selection task than on others seems dubious, as do its claims about the rarity assumption. The status of the model was also questioned. Is it a normative (see Evans and Over, 1996a&b) or a psychological model? One of the problems with Oaksford and Chater's model which received little attention in Chapter 1 is its failure to account for the importance of representation to subjects' behaviour on the selection task. For example, in recent years work has appeared which emphasises the importance of what is represented to people's patterns of performance on a variety of reasoning tasks (Girotto, Johnson-Laird and Legrenzi, 1993; Love and Kessler, 1995; Handley, 1996). Oaksford and Chater fail to consider the importance of selective representation to the process of selecting evidence for hypothesis testing. In this sense, more than any other, their model may be thought of as a normative one.

The constrained rationality model, on the other hand, is an account of how representational constraints interact with subjects' background knowledge about subjective probability in order to produce the observed patterns of information selection on the task used in Experiment 2. Underlying this account is Sperber, Cara and Girotto's (1995) assumption that people consider possible cognitive effects in the order of their accessibility and cease to consider these effects when the expected level of relevance is achieved. Accordingly, even though the model claims that the information which subjects select to
test their hypotheses is determined by both background knowledge of subjective probability and a consideration of the utility of the various outcomes of a test, the perceived utility of certain outcomes was claimed to be dependent on the manner in which subjects represent the problem. Thus, the model presented in Chapter 2 constitutes an advance on the almost wholly normative model of Oaksford and Chater.

In addition to the advance which the model constitutes, the studies upon which it is based also contain novel findings. Firstly, the basic finding of an effect of subjective probability on subjects' information selection in the task used in Experiment 2 is the first such effect found outside Wason's selection task. Such a finding also raises questions about current explanations of subjects' performance on the task. Both Mynatt, Doherty and Dragan (1993) and Evans and Over (1996b) claim that subjects are predisposed towards selecting further evidence relevant to the hypothesis which they currently consider to be most likely. The finding in Experiment 2, however, was that subjects given a strong initial piece of information (concerning a rare feature) were less likely to choose further information about the hypothesis to which that initial evidence related than were subjects given a much weaker initial piece of information (concerning a common feature). Thus, the subjects who might have been expected to have the highest level of confidence in the focal hypothesis were less likely to choose further information about that hypothesis.

It is clear that such a finding directly contradicts Mynatt et al's explanation of previous results on the task. However, the account of Evans and Over differs somewhat from that of Mynatt et al. This is because although Evans and Over do claim that subjects are likely to build a mental model of the experimental situation around the hypothesis which they currently consider to be most likely, these authors also claim that aspects of the experimental materials may lead subjects to construct a model of the experimental situation around the feature which the initial piece of evidence concerns. Thus, Evans and Over might claim that subjects given an initial piece of information concerning a rare feature tended to select further information relevant to the alternative hypothesis because their model of the experimental situation was constructed around the rare feature. The question, of course, is how sensitive subjects are to increases in the information to be
gained from certain information selections due to the rarity of the features which those selections concern. Is it simply the case that subjects' attention is focused onto the rare feature because of its rarity or do they explicitly consider the information to be gained from the selection of further evidence concerning that rare feature? The answer to this question has a bearing, not only on our conception of human abilities in relation to hypothesis testing, but also on our general view of human cognition. Unfortunately, neither the experiments described in this thesis, nor any studies currently in the literature, are of any help in answering this question.

6.3 A SUMMARY AND INITIAL DISCUSSION OF THE EXPERIMENTS ON INFORMATION USE

Whilst Experiments 1 and 2 examined subjects' selection of information for hypothesis testing, Experiments 3 - 8 examined how subjects revised their beliefs in the light of evidence relevant to competing hypotheses. These experiments can be neatly categorised into two groups. Experiments 3 - 5 followed up on the result of Experiment 1 and sought to investigate the role of expectations from background knowledge in the belief revision process. Experiments 6 - 8, on the other hand, demonstrated that there are several sources of information available to subjects when determining the strength of a piece of evidence. Experiments 6 and 7 revealed that which of these sources of information is used depends on context, whereas Experiment 8 demonstrated the sequential use of more than one source of information within the same task. The source of information used by subjects in assigning a subjective value to the evidence which they received was found to determine the occurrence or non-occurrence of an order effect. Both sets of studies were argued to provide evidence of the insufficiency of existing accounts of belief revision. In the following sections these studies will be described and discussed in more detail.

6.3.1 Experiments 3, 4 and 5: The role of expectation in belief revision

Experiments 3 - 5 were designed to investigate the unexpected result of Experiment 1. Although the information selection aspect of that study has already been discussed in the
previous section of this chapter, subjects' use of the information which they received has, thus far, been neglected. In Experiment 1, as well as choosing information to aid them decide between two alternatives, subjects were also asked to state their confidence in those alternatives in the light of the information which they received. The finding of interest from that experiment was contained in the confidence ratings of subjects who chose a further piece of evidence relevant to the hypothesis about which they already possessed some information. Of greatest interest were the confidence ratings of subjects who received evidence about a common or a rare feature and chose a second piece of evidence which also suggested that the focal hypothesis was the case. The confidence ratings of subjects in the Rare condition were significantly lower than the confidence ratings of subjects in the Common condition. This result was surprising as, on the face of it, those subjects given initial information concerning a rare feature possessed stronger evidence than the subjects who had received a first piece of evidence about a common feature. It was argued that this result was due to very high expectations about the second piece of evidence amongst those subjects in the Rare condition. The second piece of evidence failed to meet those expectations causing them to produce lower confidence ratings than subjects in the Common condition who did not have such high expectations about the second piece of evidence.

Experiment 3 was an attempt to test this expectation account of the results of Experiment 1. It was reasoned that if subjects' expectations underlay their behaviour in Experiment 1 then a recency effect should be observed when two groups of subjects were given exactly the same evidence - consisting of one strong, and one weak, evidential item - but in different orders. Such a recency effect was observed, with subjects who received the strong piece of evidence first producing significantly lower confidence ratings than were produced by subjects who received the strong piece of evidence second. Once again, it was argued that subjects who received the strong piece of evidence first had higher expectations about the second piece of evidence than did subjects who received the weaker piece of evidence first (there was a significant difference in the expected direction between the mean expectations of subjects in both groups concerning the second piece of evidence).
However, the literature on order effects, reviewed in Chapter 3, suggested an alternative explanation for these results. This alternative explanation comes from the work of Hogarth and Einhorn (1992) who have proposed that, with certain kinds of belief updating task, sequential evidence with contrasting strength will always result in recency. According to their model, weak evidence followed by strong evidence always results in upwards revision of belief, whereas strong evidence followed by weak evidence always results in downwards revision. Their model makes these predictions because it assumes that the weight assigned to new evidence will be either proportional, or inversely proportional, to the anchor (i.e. subjects' beliefs based on earlier evidence) dependent on whether that new evidence is more or less positive than background belief. Thus, when new evidence is more positive than background belief and the anchor is small, that new evidence will produce a large upwards belief revision. Conversely, when new evidence is less positive than background belief and the anchor is large, the new evidence will produce a large downwards revision. Hogarth and Einhorn also claim that the size of the recency effect is directly related to the degree of contrast between the pieces of evidence which subjects receive.

This account is firmly in the tradition of belief revision as information integration best exemplified by the work of Anderson (1962; 1965; 1981) and Shanteau (1970; 1972). According to this tradition the processes involved in belief revision may be modelled mathematically in terms of the value which subjects assign to pieces of evidence and the algebra-like operations which they carry out on those values. This approach, and associated work, was extensively reviewed in Chapter 3, as were problems with the approach. The problem most relevant to this summary is that of the independence of evidential items. That is, Hogarth and Einhorn (and many other workers in the field) assume that the values assigned to pieces of evidence by subjects are derived independently of each other. This is in sharp contrast to the account, given above, implicating expectation in the results of Experiments 1 and 3. Such an account claims that, in many cases, values will be assigned to pieces of evidence in the light of expectations caused by earlier occurring evidence.
Experiments 4 and 5 were attempts to discriminate between expectation and contrast based accounts of the result of Experiment 1 and the recency effect in Experiment 3. In Experiment 4 degree of contrast between the evidential items was controlled for by means of a pre-test. Thus, for each problem content, two sets of features were chosen. Each set consisted of one rare and one common feature. Each set was received by two groups of subjects, with one group receiving the strong evidence first and the other group receiving the strong evidence second. The crucial difference between these sets of features was the degree to which possession of the common feature was perceived to be predicted by possession of the rare feature. For one of the feature sets the common feature was strongly predicted by the rare feature (the Strong Association condition), whereas in the second set the predictive relationship was weak (the Low Association condition). Because degree of contrast was held constant across both feature sets, Hogarth and Einhorn's contrast assumption predicted recency with both pairs of evidential items. An expectation based account, on the other hand, predicted recency with the High Association set but not with the Low Association set. As predicted, recency was found for the High, but not for the Low, Association features.

Experiment 5 also used an order manipulation, where the degree of contrast across two sets of features was kept constant, to discriminate between the expectation and contrast accounts for the results of Experiments 1 and 3. In Experiments 3 and 4 all subjects were told that the rare feature was possessed by 95% of instances of the focal category, whereas 70% of instances of that category possessed the common feature. In Experiment 5 subjects were told either that 95% of instances of the focal category possessed both a common and a rare feature or, that 70% of instances of the focal category possessed both of those features. In this way, although the absolute value of the evidence given to subjects was either increased or decreased relative to the evidence received by subjects in Experiment 3 and 4, the degree of contrast between the pieces of evidence was kept constant. In both cases, there was no difference between the likelihoods that members of the focal category possessed the features. The only difference between the evidential items was the perceived rarity of the features themselves.
Once again, Hogarth and Einhorn's contrast assumption predicted recency in both the 95% and 70% conditions. The expectation based account, on the other hand, predicted recency only in the 95% condition. This prediction was based on the degree to which the expectations of subjects who received the strong evidence first would be exceeded by the second piece of evidence in both conditions. Although there was very little difference predicted between the expectations about the rare feature of subjects told that 95% of instances of the focal category possessed the common feature and the expectations of subjects told that 70% of instances possessed the common feature, the second pieces of evidence which these subjects received were markedly different. Subjects told that 95% of instances of the focal category possessed the common feature were also told that 95% of instances of that category possessed the rare feature. Subjects initially told that 70% of instances of the focal category possessed the common feature were told that 70% of instances possessed the rare feature. Thus, a greater recency effect was predicted amongst the results of subjects in the 95% condition than amongst the results of subjects in the 70% condition. Once again, as predicted, the results of Experiment 5 displayed significant recency in the 95% condition and a slight trend towards primacy in the 70% condition.

The novelty of the results described above lies in their demonstration that, at least in certain cases, the value assigned by subjects to a piece of evidence is dependent on earlier occurring evidence. These results are at odds with the account of evidence understanding implicit in most accounts of belief revision. Although Asch (1946), who provided one of the first demonstrations of an order effect in personality impression formation, accounted for his results in terms of a change of meaning hypothesis, the demonstration of context effects which cannot be explained in terms of a cognitive algebraic model is notoriously difficult. The experiments summarised here consist of such a demonstration and suggest that although subjects are performing some type of simple algebra-like operation upon the values which they assign to the evidence they receive, it is the process of value-assignment which is of most interest. In the discussion of Experiments 3 - 5 it was suggested that the process by which subjects assign values to the pieces of evidence which they receive is an inferential one. By this it is meant that subjects have expectations concerning the evidence which they receive. These expectations act as the
This suggestion, that the process by which subjective values are assigned to pieces of evidence is inferential, has several interesting implications. Firstly, it resolves the problem of how a piece of evidence is understood as being marked for one of the available candidate hypotheses. Both Lopes (1987) and Hogarth and Einhorn (1992) have pointed out how this part of the process of evidence understanding contains the possibility for error. Neither author (or set of authors) details a clear proposition of how this part of the process is achieved. The claim that it is inferential solves the specific problem of how a decision is reached about which hypothesis a piece of evidence is marked for. Thus, if the conclusion of a deductive inference involving an expectation and a piece of evidence agrees with the expectation derived, either from background knowledge or early occurring evidence and background knowledge, then the evidence supports the focal hypothesis. If the conclusion does not agree with the expectation then the evidence is marked for the alternative hypothesis.

The second interesting implication of the experiments summarised in this section of the chapter is that they suggest a view of the belief revision process which is broadly in line with current models of other cognitive phenomena. Central to this view is that no single piece of information is interpreted in isolation and that meaning is dependent on context. This context is comprised of both background knowledge and information already present in the specific situation. Such a view is to be seen in work on language. For example, Relevance theory (Sperber and Wilson, 1986; 1995) and Sanford and Garrod's (1981) model of text comprehension are both inferentially driven. Two very good specific examples are Moxey and Sanford's (1993) work on quantifiers and Shanon's (1988) analysis of similarity judgements described in Chapter 3. In the former example, the approximate quantity communicated by a quantifier is seen as being dependent on subjects' expectations about the context in which the quantifier is used. Likewise, Shanon's analysis of similarity judgements rests on the assumption that subjects will often use their background knowledge about the likelihood that a relationship of similarity exists between
two objects and then go on to search for features across which to compute actual similarity. In a similar vein, Experiments 1, 3, 4 and 5 have demonstrated how pieces of evidence may be dependent. Although the general approach taken in these experiments to the derivation of subjective values for pieces of evidence is not new, the demonstration of order effects which may only be explained in these terms is a novel one and suggests that current models of the belief revision process will, at the very least, have to be revised.

6.3.2 Experiments 6, 7 and 8: Assigning values to evidence

The experiments summarised in the previous section examined the specific question of how expectations derived from early occurring evidence are used in the interpretation of later evidence. The experiments to be summarised, and discussed, in this section were designed to examine the differential use of information, other than information about the extent to which the evidential features are associated, in subjects' assignment of values to the pieces of evidence which they received. In all of the experiments concerning belief revision contained in this thesis, subjects received evidence of the form:

A% of instances of category X possess feature B.

Apart from the extent to which possession of one feature is predicted by possession of another, each piece of evidence contained two parts about which the subject might be expected to possess some knowledge: features and likelihood expressions. Although it is obvious that subjects must integrate both of these sources of knowledge in forming an initial representation of the meaning of the piece of evidence, Experiments 6 - 8 attempted to test the intuition that subjects would focus upon these sources of information to a greater, or lesser, extent when assigning subjective values to pieces of evidence.

Experiment 6 was an exploratory investigation of this intuition where, as in previous experiments, all subjects received one strong, and one weak, piece of evidence. As before, the order in which subjects received the evidence was manipulated. Of most
interest, however, were the features about which subjects received information. Half of the
subjects were given evidence concerning two rare features whilst the remaining subjects
received evidence concerning two common features. The rarity of the features within each
set was kept approximately equal by means of a pre-test as was the perceived association
between the features which was low. Accordingly, the order manipulation was achieved
simply by telling subjects in one condition that 95% of instances of the focal category
possessed the first feature and that 70% of instances possessed the second feature. Subjects
in the second order condition, on the other hand, were told that 70% of instances possessed
the first feature and 95% of instances possessed the second feature.

Some cautious predictions were made about the results of this experiment. It was
predicted that whilst assigning subjective values to the pieces of evidence which they had
received, subjects would attend to different elements of these pieces of evidence dependent
on the rarity of the features which they concerned. Thus, subjects who received evidence
about two rare features were expected to focus on the features (because of their rarity)
whilst subjects who received evidence about two common features were expected to focus
on the likelihood element of the information which they received. This latter expectation,
of reduced focus on the features amongst subjects who received evidence concerning
common features, was based on the intuition that common features would be less
informative, and therefore less relevant, to subjects than would rare features. The effect of
this difference in the information upon which subjects focused was expected to manifest
itself in the presence, or absence, of order effects. That is, the results of subjects who
received information about common features were expected to contain an order effect,
whereas the results of subjects given information about rare features were not expected to
display an effect of information order. That part of the evidence which subjects focus upon
should determine the degree of contrast between the pieces of evidence which they receive.
Thus, if subjects focus on the features there should be little contrast between the pieces of
evidence as both features are of approximately equal rarity. Conversely, if subjects focus
on the likelihoods there should be greater contrast between the pieces of evidence as there
is a difference between the likelihoods (95% vs. 70%).
Interestingly, most models of belief revision (e.g. Anderson, 1981; Lopes, 1987; Hogarth and Einhorn, 1992) make the same prediction. However, they are only capable of doing so when the degree of contrast between the pieces of evidence is known in advance. The (weak) predictions made at the outset of this experiment were made in the absence of such knowledge. Rather, they were made on the basis of which element of the pieces of evidence it was felt would be most informative to the subjects. Although the result of interest in this experiment was not significant, the relevant data plot suggested a trend towards recency with common features but no such trend was apparent with rare features.

Experiment 7 followed up on the results of Experiment 6. Once again an order manipulation was employed with all subjects being told that 95% of instances of the focal category possessed one feature and 70% of instances possessed the other. Unlike the previous experiments, however, the problem contents used in Experiment 7 were abstract, with subjects being asked to categorise parts of speech from an ancient form of an imaginary language based on what is known about the modern form of that language. The results of Experiment 7 replicated, almost exactly, the results of those subjects in Experiment 6 who had been given information about common features. It was argued that as subjects in Experiment 7 could not have been assigning values to the pieces of evidence which they received based upon background knowledge about the features, the results of Experiment 7 constituted confirmation of the claim that subjects in Experiment 6, who received evidence concerning common features of the objects to be categorised, did not use information about those features from background knowledge to assign subjective values to the pieces of evidence which they received.

The results of Experiments 6 and 7 are both interesting, and novel, for several reasons. Firstly, no existing account of belief revision could have predicted the similarities present in the results of Experiments 6 and 7 in advance. Once again, this is because the emphasis in existing models of the belief revision process has been on the integration of evidence rather than on the process of assigning a subjective value to new information. It should be pointed out, however, that there are several accounts of belief revision in existence which can account for the results of Experiments 6 and 7 in a post hoc manner.
Such an explanation would rely on the simple notion of averaging (or contrast in the case of Hogarth and Einhorn, whose model accounts for averaging-like behaviour). Nonetheless, in order to predict when an order effect will, or will not, occur, it is necessary to be capable of predicting when there will be contrast between the pieces of evidence which subjects receive. Current model of belief revision are incapable of making such predictions.

The similarity which was found between the confidence ratings of subjects given information about common features in Experiment 6 and the confidence ratings produced by subjects in Experiment 7 were predicted on the basis of what was felt would be relevant to subjects engaging in the task. Thus, subjects given information about rare features were expected to focus on the features when assigning subjective values to the evidence which they received. Subjects given information about common, and abstract, features were expected to focus on the likelihoods associated with those features, rather than on the features themselves. The confirmation of these predictions is interpreted as providing support for the view that human cognition is bound by Sperber, Cara and Girotto's (1995) effort/effect trade off. This echoes the interpretation which was given of the results of Experiment 2 where subjects' behaviour was explained in terms of an interaction between background knowledge and a limited representational system. Thus, those subjects in Experiment 6 who were given information about common features assigned subjective values to the pieces of information which they received on the basis of the most informative component of the pieces of evidence which they received. As the features were common, they were also relatively uninformative, and focusing on the likelihoods with which those features were possessed by instances of the focal category represented the strategy likely to lead to most cognitive effects in terms of deciding between the hypotheses.

What is most interesting about this account of the results of Experiments 6 and 7 is that it suggests an alternative approach to the study of decision making. Previous work has tended to focus on subjects' non-normative use of base rates (e.g. Tversky and Kahneman, 1980; Bar-Hillel, 1980) or likelihoods (e.g. Beyth-Marom and Fischhoff, 1983; Ofir,
Indeed, even the more recent work of Gigerenzer and Hoffrage (1995) and Cosmides and Tooby (1996) which claims to demonstrate the normative use of base rate information by subjects (when presented in terms of frequencies) might be said to share the view that base rates, or likelihoods, should be viewed as basic units of information. Whilst this may be correct in a Bayesian sense, it is clearly inappropriate from a psychological point of view. Although some work does exist (described in Chapter 3) which suggests that the information which subjects attend to in a decision making situation is context-dependent (e.g. Macchi, 1995; for a discussion of this topic also see Evans and Over, 1996b) this work still regards base rates as the basic unit of analysis. Experiments 6 and 7 in this thesis, on the other hand, suggest that a more thorough analysis is required. A base rate, for example, consists of both a feature and a probability expression. In certain circumstances the information contained in the base rate will not be treated by subjects as a unit. Thus, in certain situations, subjects may concentrate on the probability expression (for example, when the feature for which the base rate is being given is uninformative), whilst in others they may focus on the feature itself.

That subjects are very sophisticated users of these different sources of information is demonstrated by the results of Experiment 8. In that experiment all subjects received a likelihood ratio in sequence. That is, all subjects were initially told the percentage of instances of one category which possessed a certain feature and were then told the percentage of instances of a second category which possessed the same feature. Two different likelihood ratios were used as well as two different features (one rare and one common). The results of this study revealed significant differences, due to feature rarity, between the initial confidence ratings produced by subjects but, in general, not between their final confidence ratings. Significant differences in final confidence ratings were due to the different likelihood ratios which subjects received.

These results are interesting for several reasons. Firstly, they reveal that subjects may switch their focus from the feature component of the pieces of evidence onto the likelihood component within the same task. It was argued that the significant differences between subjects' initial confidence ratings was due to the subjects who received
information about a rare feature using their background knowledge about that feature's rarity when assigning a subjective value to the piece of evidence which they received.

Conversely, subjects who received an initial piece of information concerning a common feature were claimed to have focused more on the likelihood expression when deriving a subjective value. However, as the second piece of evidence which subjects received always consisted of the second half of a likelihood ratio, subjects in all conditions focused on the value of that ratio rather than on the feature for which the ratio had been given. In terms of discriminating between the hypotheses, the likelihood ratio was the most informative information upon which to base their final confidence ratings.

A second point of interest in the results of Experiment 8 is their implications for current models of information integration, generally, and belief revision in particular. The effect of information order was to create a greater difference between the initial confidence ratings of subjects given information about a rare feature than between the initial confidence ratings of subjects given an initial piece of information concerning a common feature. According to Hogarth and Einhorn's contrast assumption this should have resulted in a greater recency effect amongst the final confidence ratings of subjects given information about rare features than amongst the final ratings of subjects given information about common features. This was not the case. In both halves of the experiment, the size of the trend towards recency was equal. In terms of general approaches to information integration, these results are interesting because they cast doubt on the claim that the interpretation of a piece of evidence is irreversible. Most models of information integration (see Massaro and Friedman, 1990 for a detailed review) assume that once a piece of evidence has been encoded and integrated, it cannot be re-encoded and re-interpreted by the subject. The results of Experiment 8 suggest that subjects who received information about rare features were able to re-interpret the initial piece of evidence which they received. Although these subjects seemed to be using background knowledge about subjective probability when assigning a subjective value to the initial piece of evidence which they received, their final confidence ratings seem to have been based on the likelihood ratio with no effect of feature rarity. Accordingly, these subjects seem to have re-interpreted the initial piece of evidence which they received in the light of later
evidence. The final point of interest in these results is how closely the final confidence ratings of all subjects resembled those which would be prescribed by Bayes theorem. Although subjects' normative accuracy in their use of the likelihood ratio which they received was not the focus of Experiment 8, nevertheless, the experiment does provide evidence that subjects are capable of using likelihood information accurately. However, it was not claimed that the process by which subjects arrived at their final confidence ratings involved the use of Bayes theorem.

Perhaps the most novel aspect of the findings from all three of the experiments summarised in this section is their suggestion that subjects focus on different aspects of the information which they receive in a judgement task, and that the locus of their attention is dependent on considerations of cognitive effect. Experiment 8, in suggesting that the locus of subjects' attention can switch in the course of the task, also suggests that subjects construct a model of the judgmental situation which contains both new and old information. The information which they attend to within this model is that which is most likely to have cognitive effects in terms of deciding between the alternatives.

6.4 THE ROLE OF BACKGROUND KNOWLEDGE IN INFORMATION SELECTION AND BELIEF REVISION

In many ways this thesis might be said to be almost exclusively concerned with the role of background knowledge in the selection and use of information for hypothesis evaluation. However, because the literatures pertaining to selection and use are at different stages of development, the experiments contained in this thesis relate to these separate literatures in different ways. Broadly speaking, the experiments on information selection may be seen as contributing to a debate currently taking place in the selection task literature. On the other hand, it is hoped that the experiments on information use will initiate such a debate in the belief revision literature. This section of the chapter will outline the main contributions of the experiments described in this thesis both to what is known about the role of background knowledge in information selection and use, and to
the debate of how best to characterise background knowledge when considering information selection and information use.

It is clear from all of the experiments presented in this thesis, that background knowledge plays a role in the selection, and use, of information for hypothesis evaluation. Much of the work in the reasoning literature has been concerned with the exact specification of this role. Part of the novelty of the findings of Experiments 1 - 8 is their demonstration of a role for background knowledge in the pseudodiagnosticity task (or variants of that task). Three sources of background knowledge have been shown to affect the way in which subjects select and use information on the task. These are: background beliefs about the rarity of the features; information about the degree to which possession of one feature is predicted by possession of another; and information about the use of probability expressions. The first of these sources of information was found to play a role in information selection and it is the role and characterisation of background knowledge in selection tasks which will be discussed first.

In Chapter 1 three separate accounts of how background knowledge influences the selection of information for hypothesis testing were discussed. The first of these was Cheng and Holyoak's (1985) pragmatic reasoning schemas account of Wason's selection task. Essentially, their account claims that subjects' information selection is guided by the activation of rules contained in schemas. These schemas are said to be induced from "ordinary life experiences" such as permissions and obligations and are defined in terms of goals and relationships to these goals. These goals constrain the inferences which subjects make to those which are pragmatically useful.

A second account of how background knowledge affects subjects' patterns of information selection has been given by Sperber, Cara and Girotto (1995). These authors take a much more flexible view of the role of background knowledge in information selection and suggest that its operation is guided by a trade-off between cognitive effort and cognitive effect. Thus, Sperber et al argue that subjects will consider those pieces of information which are likely to yield the greatest cognitive effects (the greatest number of
inferences and new beliefs etc.) for the least amount of cognitive processing. In Chapter I this general approach was contrasted with the approach of Oaksford and Chater (1994) who argued that subjects calculate the expected information to be gained from each possible piece of information and decide which pieces to select based on these calculations (for a discussion of the decision-theoretic approach to information selection, see the next section). These calculations are based on the subjects' background knowledge of the probability of events in their environment.

It was argued in Chapter I that Cheng and Holyoak’s (1985) characterisation of the structure of background knowledge was too rigid to be of general value. Indeed, it is difficult to see how the pragmatic reasoning schemas approach might usefully be applied to the results of the experiments in Chapter 2 where the pseudodiagnosticity task was used. It does not seem possible to argue that subjects were applying rules contained in some general purpose schema when deciding which evidence to select. Whilst it is possible that some general strategy exists for selecting information with which to test hypotheses (see Klayman and Ha, 1987), it is unlikely that this strategy is captured, in every case, by schemas.

Cheng and Holyoak’s approach rests on the claim that the general purpose schemas which they claim to exist are domain specific. As was pointed out in Chapter 1, both Oaksford and Chater (1993) and Johnson-Laird and Byrne (1991) have questioned the validity of the schema approach by pointing out that no general principles have yet been identified which specify how a domain is to be isolated. The work of Wattenmaker (1995) - described in Chapter 3 - is very interesting in this regard. He has speculated that knowledge may be compartmentalised into domains on the basis of category structure. He claims that the factor underlying this compartmentalisation is whether categorisation can take place on the basis of summing the features of the object to be categorised. This seems to be the case for social, but not for object, knowledge. This approach to the question of domain specificity deals with knowledge about objects rather than knowledge about the relationship between objects. Nor does it deal with the application of rules. Accordingly, it
is likely to much more flexible than the approach favoured by Cheng and Holyoak. It also has the advantage of being a principled distinction between domains.

Wattenmaker's approach to domain specificity is also more compatible with the pseudodiagnosticity task. Based on the experiments reported in this thesis, it would seem that the pseudodiagnosticity task is best viewed as a task of categorisation rather than as a reasoning task. Subjects are asked to select information in order to help them categorise an object. Interestingly, in selecting, and using that information subjects have been shown to be sensitive to both feature frequency and the degree of association between features. As was discussed in Chapter 3, these are factors which are known to underlie the learning of category structure. In fact, one of the advantages of the pseudodiagnosticity task as an instrument for the investigation of the factors underlying information selection for hypothesis testing is that it is not amenable to a pragmatic reasoning schemas interpretation. The knowledge activated in the course of the task cannot be characterised as possessing a schema-like structure. Instead, the task allows for an unconfounded examination of how knowledge about the features of objects is used when selecting information for the purposes of deciding between hypotheses. In terms of domain specificity, the task might be used to see whether object and social knowledge have different effects on the selection, and use, of evidence for hypothesis evaluation.

In Chapter 1 it was argued that a combination of the approaches of Oaksford and Chater and Sperber, Cara and Girotto to the question of how background knowledge is used when selecting information to test hypotheses was likely to be most successful. Certainly, in terms of how this background knowledge is characterised, Oaksford and Chater seem to have more to say that is relevant to the pseudodiagnosticity task than do Cheng and Holyoak (1985). Their account is cast at the level of the probabilities of features of objects which is very similar to the level at which the constrained rationality model, outlined in Chapter 2, works. As was pointed out in Chapter 2, however, one of the theoretical responsibilities of any model of how background knowledge operates to guide information selection is to specify how its operation is constrained. Humans have limited representational resources so the argument that they calculate the information which they
expect to gain from the selection of each possible piece of evidence is untenable. Although, as was argued in Chapter 1, Sperber et al fail to give an account of the specific sources of information available to subjects in any given context, they do specify a principle by which the activation of background knowledge might be specified in a hypothesis evaluation task. The constrained rationality model relies on this principle, of a trade-off between effort and effect, in order to capture the data from Experiment 2.

As was mentioned in the opening paragraph of this section, the literature on belief revision is at a very different stage of development to the literature on reasoning and information selection. As has just been seen, the information selection literature is relatively sophisticated in its attempts to capture the effects of background knowledge on the processes involved in choosing evidence with which to evaluate hypotheses. The belief revision literature, on the other hand, is striking in its failure to consider background knowledge as a factor in the use of evidence. Experiments 3-8 clearly demonstrate that background knowledge about features of objects plays a role in subjects' use of information about those objects for purposes of hypothesis evaluation. Experiments 3-5 demonstrate the importance of degree of association between features to the process of assigning subjective values to pieces of evidence. Experiments 6-8 demonstrate the role of selective attention, guided by background knowledge, to the occurrence (or non-occurrence) of an order effect. In so doing, all of these experiments emphasise elements of the belief revision process which have received very little attention in the experimental literature.

It was pointed out in Chapter 3 that existing accounts of belief revision (both algebraic and procedural) rely on the independence assumption for their operation. That is, these accounts assume that the subjective values of pieces of evidence (even when they occur in sequence) are derived independently of each other. Thus, when Hogarth and Einhorn (1992) claim that the contrast between the subjective value of pieces of sequential evidence underlies recency, they are claiming that, given a subjective value for one piece of information, if the subjective value of later evidence is much greater, or much less, than the value of this earlier evidence, recency will occur. Strikingly, Hogarth and Einhorn fail
to account for the process by which the subjective value for any piece of evidence is
derived. It was argued in Chapter 3 that there are several a priori reasons for doubting the
assumption that these values are derived independently. Wattenmaker's (1995) work on
domain specific differences in knowledge organisation was cited, as well as work on
feature correlation in the concept literature and more general work (e.g. Shanon, 1988) on
representational theories of meaning.

Experiments 1, 3, 4 and 5 also suggest that the independence assumption is
unwarranted. Their results suggest that, in certain cases, early evidence gives rise to
expectations about later occurring evidence (these expectations are due to background
knowledge about the degree of association between the evidential features). These
expectations determine the subjective value assigned to the later evidence. Experiment 4,
in particular, demonstrates that the occurrence of a recency effect is dependent on these
expectations from background knowledge. In Experiment 4 subjects received information
about two features which were either high, or low, in association. Recency was only
observed amongst the results of those subjects given information about high association
features. This finding was explained in terms of subjects in the high association condition
being given early evidence which led to very strong expectations about subsequent
evidence. Because the subsequent evidence did not meet those expectations, these subjects
left their initial confidence ratings unchanged. Subjects who received the same evidence,
but in a different order, did not have such high expectations. Accordingly, their final
confidence ratings were significantly higher than both their initial confidence ratings and
the final confidence ratings of subjects who had received the evidence in the reverse order.
In the low association conditions, no such recency effect was observed.

This result suggests that both the independence, and contrast, assumptions are
inadequate. Background knowledge (about the association between evidential features, in
the case just described) has a role to play in the interpretation of evidence. For this reason,
it is not enough to rely on the notion of contrast between the subjective values of pieces of
evidence to explain recency. An account is also needed of the circumstances under which
contrast will occur and background knowledge is likely to be central to any such account.
The results of Experiments 6 - 8 demonstrate that there are determinants of contrast other than background knowledge about the degree to which evidential features are associated. The finding of a trend towards recency with common features in Experiment 6, but not with rare features, suggests that subjects were focusing on the features in the latter case and on the likelihood expressions in the former. The similarity of the results of Experiment 7, where abstract materials were used, to the results with common features in Experiment 6, also strengthens this suggestion. Contrast will exist between two pieces of evidence only if they differ on the dimension upon which subjects tend to focus. It was argued that subjects in the common features condition of Experiment 6 focused on the likelihood expressions. As the pieces of evidence which they received differed along this dimension, there was contrast between the subjective values assigned to the pieces of evidence and recency was observed. That subjects were focusing on the likelihood expressions is also suggested by the similarity of these results to the results of Experiment 7 where subjects had no background knowledge about the features used.

Interestingly, the suggestion that subjects were focusing on the likelihood component of the evidence which they received in the common condition of Experiment 6 and in Experiment 7, suggests that subjects were using information about likelihood expressions when assigning subjective values to the pieces of evidence which they received. This suggests that it is not possible to give an account of information interpretation, and use, wholly in terms of subjects' background knowledge about the probabilities of the features they receive evidence about. Likelihood expressions are quantifiers, however, and existing accounts of the operation of linguistic quantifiers (Moxey and Sanford, 1993) might be expected to generalise to the operation of likelihood expressions.

The results of Experiments 3 - 8 also suggest that the manner in which the pseudodiagnosticity task was characterised, earlier in this section, was accurate. It was argued that the task is one of selecting information in order to categorise an object. In this sense, many of the factors found to affect both the learning of categories, and
categorisation judgements, should also be found to affect judgements based on information received in the pseudodiagnosticity task. This has been found to be the case with both association between features (Experiments 1, 3, 4 and 5), and feature frequency (Experiments 6 - 8), proving to underlie subjects' assignment of subjective values to the evidence which they received. The finding of an effect of feature association is particularly interesting in terms of information selection. Although this thesis does not provide evidence that subjects' background knowledge about the degree to which possession of one feature is predicted by possession of another affects the information which they select to test hypotheses, it is likely that such effects will be found with the appropriate stimulus materials. If subjects are sensitive to such associations when assigning subjective values to pieces of information, then it is highly likely that they will also take this information into account when considering the probability of various test outcomes given information about an associated feature.

6.4 WHAT DO INFORMATION SELECTION AND INFORMATION USE HAVE IN COMMON?

In the introduction to this thesis it was pointed out that there have been several recent attempts to integrate the fields of reasoning, and decision making, research. For example, the work of Johnson-Laird (1994) and Evans, Manktelow and Over (1993) was cited. The former has attempted to apply principles derived from the study of deductive reasoning to probabilistic thinking, whereas the latter workers have attempted to recast Wason's selection task (traditionally thought of as a reasoning problem) in terms of decision making. In addition to the links made in the previous section, which concerned background knowledge, this section of Chapter 6 will attempt to enlarge upon links already known to exist between these (supposedly) different processes, and, hopefully, to make new links.

The most obvious place to start is in terms of the experimental tasks which subjects were asked to perform in the course of this thesis. In Experiments 1 and 2 subjects were
asked to choose one of three possible pieces of evidence in order to help them decide
between two alternatives. Also in Experiment 1, and in Experiments 3 - 8, subjects were
given some evidence and asked to rate their confidence in a pair of hypotheses in the light
of that information. The distinction between the two types of task may be summarised as
follows: The choice task involves a decision about what is likely to constitute good
evidence whilst the rating task involves deciding how good some evidence, already
received, actually is. Stated in these terms the tasks appear to be very similar. However, as
a reading of the literature reviews contained in Chapters 1 and 3 reveals, the tasks of
information choice and information use have received vastly different treatments in their
respective literatures. The next section will constitute a discussion of the extent to which
the pseudodiagnosticity task, and information selection tasks in general, may be
successfully viewed as decision making tasks. The penultimate section of the chapter will
discuss the role played by inference in tasks traditionally thought of as involving
judgement or decision making.

6.4.1 Information selection as decision making

One of the themes of Chapter 1 was the role played by general considerations of
relevance and specific considerations of subjective probability in the selection of
information for hypothesis testing. It was seen that a tension exists within the literature on
information selection between workers such as Sperber, Cara and Girotto (1995) who wish
to account for information selection in terms of the general cognitive principles of effect
and effort, and other workers (Oaksford and Chater, 1994; 1995) who claim that subjects'
background knowledge of, and assumptions about, subjective probability provide a means
of quantifying intuitions about relevance. On one level, there is a very fundamental
difference between these two approaches as one views the subject as being involved in an
act of communication whilst the other views the subject as a scientist attempting to decide
which is the best experiment to carry out. On another level, of course, the distinction
between these two approaches is trivial as both are agreed that a complete understanding of
subjects' pattern of information selections in hypothesis testing tasks may only be
understood via an understanding of the background knowledge and pragmatic assumptions which subjects bring to the task.

Despite this tension, at the time of writing there does seem to be a strong consensus within the field that information selection for hypothesis testing may be viewed, in part at least, as a behaviour which involves decision making. Thus, alongside the work of Oaksford and Chater lies the work of Kirby (1994a&b) and Evans and Over (1996a&b; Evans, Over and Manktelow, 1993). What all of these workers have in common is the belief that an analysis of the subjective probabilities of outcomes, and their associated utilities, mirrors some of the processes involved in human hypothesis testing behaviour. The constrained rationality model presented in Chapter 2 is couched in these terms. It claims that subjects make decisions when selecting evidence and that these decisions are based upon a consideration of the information which the available pieces of evidence are likely to yield.

The model, however, does not claim that subjects engage in a thorough analysis of the outcomes contingent on the choice of each piece of information available. Rather, it views subjects as performing a limited analysis. In this respect, it resembles the work of Evans and Over more closely than it does the work of Oaksford and Chater. This resemblance is to do with the recognition that the human representational system is, in general, incapable of the amount of cognitive effort required to consider all possible outcomes, and their associated utilities, when selecting information to test hypotheses. Thus, subjects will perform an analysis of those pieces of information which appear most relevant and, to paraphrase Sperber, Cara and Girotto (1995), stop when their expectations of relevance have been met. There are many possible sources of relevance, other than subjective probability. For example, Evans' (1984; 1989) work on matching bias in the abstract selection task suggests that subjects' representation of the conditional rule in the task is guided by sources of linguistic relevance. Likewise, Evans and Over (1996b) suggest that, in many cases, evidence pertaining to the hypothesis which they currently think most likely, assumes relevance for subjects. Although the results of Experiment 2
suggest that this may not always be the case, what is important is the search for principles underlying the selective attention which is so essential to subjects' behaviour on the task.

Although there is nothing new in this kind of "bounded rationality" argument, what is novel about the constrained rationality model is its demonstration of how probabilistic factors may operate within cognitive constraints to determine subjects' patterns of information selection on a hypothesis testing task. Thus, the information which subjects select to test the alternatives present in the scenario which they receive is determined both by their assignment of probabilities and utilities to potential test outcomes and their initial representation of the task. In this sense, the model presented in this thesis may be deemed to be more psychological in its concerns than is Oaksford and Chater's model, reviewed in Chapter 1. Accordingly, although patterns of information selection may, on one level, be said to be the product of decision-theoretic processes, the operation of these processes is confined to a subset of the possible evidence. This subset is determined by a trade-off between cognitive effect and effort.

6.4.2 Belief revision, inference and relevance

In the previous section, the argument that information selection may be described in decision-theoretic terms was briefly discussed. In this section it will be argued that belief revision, a type of decision making, shares many of the features traditionally associated with the reasoning literature and selection tasks in general. The traditional view of hypothesis testing is that it involves the application of the hypothetico-deductive method (Popper, 1959). That is, the hypothesis tester must infer testable consequences of the hypothesis currently under consideration. As was described in the previous section, the current consensus is that selection tasks may best be described in decision theoretic terms (with the exception of Sperber, Cara and Girotto who seem to favour the traditional approach in their characterisation of Wason's selection task). Paradoxically, one of the claims made in this thesis is that belief revision tasks, traditionally viewed in the same light as decision making, involve a large amount of inference.
This claim was first made in Chapter 3 where the differences between knowledge about objects and knowledge about social phenomena was discussed. It has been suggested (Asch and Zukier, 1984; Wattenmaker, 1995) that social knowledge is distinguished from object knowledge by its largely inferential nature. In other words, subjects are able to infer a cause for inconsistencies between pieces of social information (Asch and Zukier 1984; Kunda, Miller and Claire, 1990) whereas this is likely to be much more difficult with inconsistent information about objects. However, it was also pointed out that the very inflexibility of object knowledge necessitates the involvement of inferential processes when object knowledge is being integrated. Given some information about features of an object, a subject is likely to infer some very strong hypotheses about other features of that object which will be tested against information in the world. For example, a subject told that a piece of furniture has a seat might infer that it is a chair and hypothesise that it will also possess some legs. This hypothesis may be tested against information about whether the piece of furniture actually possesses legs.

Experiments 1, 3, 4 and 5 strongly suggest that this kind of hypothetico-deductive reasoning underlies, at least some, order effects in belief revision. It was seen that subjects, when told that a high percentage of instances of an object possessed a certain feature, generated expectations about the likelihood that instances of the same object would possess a second feature which they knew, from background knowledge, to be highly associated with the first feature. These expectations may be characterised as hypotheses which were tested against information about the actual percentage of instances of the object which possessed the second feature. It was pointed out that this kind of inference provides a solution to the problem of how subjects decide which of the alternative hypotheses, present in the scenario, the evidence is marked for. What is interesting about such a solution is that it characterises evidence interpretation in decision making and judgement tasks in inferential terms. Thus, not only does the process of evidence selection involve decision making, the process of making judgements on the basis of evidence involves inference.
The phenomena of evidence selection and belief revision are even more closely related, however, as Experiments 6 - 8 demonstrated. In the previous section the importance of selective attention to evidence selection was discussed. It was argued that subjects attend to a subset of the possible evidence and that their attention is directed by considerations of relevance. Experiments 6, 7 and 8 demonstrated that selective attention also plays a part in the process of assigning a subjective value to a piece of evidence. The results of Experiment 6 suggested that subjects attend to the most informative component of the pieces of evidence which they receive. Thus, subjects given information about rare features tend to focus on the features themselves when assigning subjective values to the evidence. On the other hand, subjects who receive information about common features focus on the likelihood associated with those features. This interpretation of Experiment 6 was confirmed by the results of Experiment 7. It is likely that selective attention plays a role in the assignment of subjective values to pieces of evidence for exactly the same reasons as it does in evidence selection. The human representational system is subject to constraints, and subjects will tend to focus on that aspect of a piece of evidence which they know from background knowledge to be most likely to discriminate between the hypotheses.

Experiment 8 also supports such an analysis of the processes involved in assigning values to evidence. In Experiment 8 the locus of subjects' attention was observed to switch from one component of the evidence (the features) to another (the associated likelihoods). This was explained, once again, in terms of the information which was most likely to discriminate between the alternatives present in the scenario. In the absence of a full likelihood ratio, subjects, given evidence concerning rare features of the object to be categorised, focused on those features when deriving a subjective value for the evidence. Once the second piece of evidence was received, and subjects possessed a complete likelihood ratio, focus switched onto the likelihoods themselves.

The results of Experiments 6 - 8 are, in some respects, very similar to results obtained by Macchi (1994) and Shafir (1993). Both of these experimenters demonstrated that subjects are more likely to base a probability estimate, or a choice, on information
which is relevant to that estimate, or choice. Thus, Shafir found that subjects were more likely to choose an option which possessed one or two very positive attributes, but that they were also more likely to reject the same option if it possessed some negative attributes. In other words, subjects will focus on that aspect of the available evidence which is most relevant to the task which they are asked to perform. When asked which of two alternatives they would accept, subjects focus on reasons for acceptance, whereas when the task is framed in terms of rejection subjects focus on reasons for rejection.

Likewise, Macchi, using the base rate problem, found that subjects’ use of the base rates which they received was dependent on both the context, and the specific questions they were asked. If the questions were framed so as to make the base rates seem relevant, then subjects used them in generating their probability estimate.

What is novel about the results of Experiments 6 - 8 is their suggestion that pragmatic processes of selective attention are at work even when subjects assign values to single pieces of evidence. As was pointed out in the summary and initial discussion of those experiments, their results suggest that the processes underlying evidence interpretation will have to be re-thought. These results also suggest that the task facing workers in the areas of both evidence selection and decision making, judgement and belief revision is the same. This task is one of specifying how a piece of evidence, an aspect of a piece of evidence, or a potential piece of evidence, assumes relevance for the subject in any context. Essentially, the task reduces to one of determining what the important sources of information are for the organism. Whilst the experiments described in this thesis suggest that knowledge about feature rarity is one such source of information, as we have seen, there are, at least, several others.

6.5 CONCLUDING REMARKS

The research contained in this thesis has extended knowledge about information selection and information use for the testing of hypotheses in several ways. It has demonstrated an effect of subjective probability on a propositional reasoning task which is
closely related to the pseudodiagnosticity paradigm. Furthermore, an account of these results has been offered suggesting that considerations of subjective probability and utility, as well as considerations of relevance, are responsible for their production. The thesis also contains a demonstration that the assumption that pieces of evidence are interpreted independently prior to their use for hypothesis evaluation is untenable. In many cases, later evidence is interpreted in the light of expectations derived from early evidence. In the absence of these expectations, the thesis demonstrates the operation of processes of selective attention in assigning values to pieces of evidence. It has been argued that the factors guiding attention in value assignment are the same as those underlying subjects' limited consideration of possible evidence in the experiments on information selection. The thesis implies that future research into information selection should investigate the extent to which subjects are aware of the importance of subjective probability in evidence selection. It also suggests that the central problem facing researchers in reasoning and in decision making is the same. This problem is one of giving a principled account of the factors underlying selective attention.
APPENDIX I - Example stimulus materials from Experiments 1 - 8

Table of Contents

Appendix 1a - Examples of the materials used in Experiment 1 .................. 301
Appendix 1b - Examples of the materials used in Experiment 2 .................. 303
Appendix 1c - Examples of the materials used in Experiment 3 .................. 307
Appendix 1d - Examples of the materials used in Experiment 4 .................. 310
Appendix 1e - Examples of stimulus materials from Experiment 6 ............. 313
Appendix 1f - Example materials from Experiment 7 ............................... 316
Appendix 1g - Example materials from Experiment 8 ............................... 319
Appendix 1a - Examples of the materials used in Experiment 1: Common Features, 65% Likelihood, Positive Feedback Condition

Engineer Problem

Your friend is an engineer and works for a large construction company. It's either company X or company Y, but you can't remember which. You do remember that he earns over £14,000 a year and that he drives a company car.

You have the following piece of information:
(a) 65% of engineers working for company X earn over £14,000 a year.

Three additional pieces of information are also available:

(b) The percentage of engineers working for company Y that earn over £14,000 a year. 25%
(c) The percentage of engineers working for company X that drive a company car. 75%
(d) The percentage of engineers working for company Y that drive a company car. 25%

Assuming that you could find out only ONE of these pieces of information (b, c, or d), which would you want in order to help you decide what company your friend works for?

Once you have marked your answer please peel back the sticker which corresponds to that answer. Underneath is the piece of information which you have selected. Now turn to the next page where you will find a scale designed to measure your confidence in each of the decision alternatives in the light of the information you have available you. One end of the scale corresponds to complete certainty that your friend works for company X and the other complete certainty that he now works for company Y. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to that alternative.

University Problem

Your neighbour has just gone away to university. You can't remember whether she attends university X or university Y. You do remember that she achieved two grade Cs and one grade B at A-level and that she has had to take up a foreign language. You have the following piece of information:
(a) 65% of students at university X achieved two grade Cs and one grade B or better at A-level.

Three additional pieces of information are available:

(b) The percentage of students at university Y that achieved two grade Cs and one B or better. 25%
(c) The percentage of students at university X that has to take up a foreign language. 75%
(d) The percentage of students at university Y that has to take up a foreign language. 25%

The remainder of the text was as for the Engineer problem. For all of the problems used in Experiment 1, the feedback (i.e. the percentages corresponding to each possible information selection), was concealed by separate opaque stickers.
Hotel Problem

Your friend is staying in a hotel in Paris. You can't remember whether it's hotel X or hotel Y. You do remember that the room costs £35 a night and that it has an en-suite bathroom. You have the following piece of information:

(a) 65% of the rooms in hotel X cost £35 a night

Three additional pieces of information are available:

(b) The percentage of rooms in hotel Y that cost £35 a night 25%
(c) The percentage of rooms in hotel X that have en-suite bathrooms 75%
(d) The percentage of rooms in hotel Y that have en-suite bathrooms 25%

The remainder of the text was as for the Engineer Problem
Appendix 1b - Examples of the materials used in Experiment 2: Common Features

Condition

Car Problem

Your sister bought a new car in 1988. You can't remember whether it's a model X or a model Y but you do remember that it has four doors and a radio.

We have already asked the following question for you and have given you the answer:

**Question A**

whether model X cars bought in 1988 have a radio

Answer: YES

Three additional questions are possible which we have listed below along with their possible answers:

**Question B**

whether model Y cars bought in 1988 have a radio

Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

**Question C**

whether model X cars bought in 1988 have four doors

Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

**Question D**

whether model Y cars bought in 1988 have four doors

Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

Assuming that you could discover the answer to only one of these questions (B, C, or D), which would you ask in order to help you decide which model car your sister drives?

Please circle your choice:  B  C  D
Engineer Problem
Your friend is a civil engineer and works for a large construction company. It's either company X or company Y, but you can't remember which. You do remember that she earns £25,000 a year and drives a company car.

We have already asked the following question for you and have given you the answer:

Question A
whether civil engineers working for company X earn £25,000 a year
Answer: YES

Three additional questions are possible which we have listed below along with their possible answers:

Question B
whether civil engineers working for company Y earn £25,000 a year
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

Question C
whether civil engineers working for company X drive company cars
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

Question D
whether civil engineers working for company Y drive company cars
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

Assuming that you could discover the answer to only one of these questions (B, C, or D), which would you ask in order to help you decide which company your friend works for?

Please circle you choice  B  C  D

304
House Problem

Your friend has just bought a new house. It's either on street X or street Y but you can't remember which. You do remember that the house has a garden and a garage.

We have already asked the following question for you and have given you the answer:

Question A
whether houses on street X have gardens
Answer: YES

Three additional questions are possible which we have listed below along with their possible answers:

Question B
whether houses on street Y have gardens
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

OF LITTLE USE
OF GREAT USE

Question C
whether houses on street X have garages
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

OF LITTLE USE
OF GREAT USE

Question D
whether houses on street Y have garages
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

OF LITTLE USE
OF GREAT USE

Assuming that you could discover the answer to only one of these questions (B, C, or D), which would you ask in order to help you decide which street your friend lives on?

Please circle your choice  B  C  D
Spanish Villa Problem

Your parents rented a villa in Spain last summer but you can’t remember whether they rented it from company X or company Y. You do remember that the villa was built sometime in the last twenty years and that it cost £150 per week.

We have already asked the following question for you and have given you the answer:

**Question A**
whether company X villas cost £150 per week
Answer: YES

Three additional questions are possible which we have listed below along with their possible answers:

**Question B**
whether company Y villas cost £150 per week
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

**Question C**
whether company X villas were built sometime in the last twenty years
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

**Question D**
whether company Y villas were built sometime in the last twenty years
Possible Answers: YES or NO

How useful would knowing the answer to this question be in deciding between the alternatives?

<table>
<thead>
<tr>
<th>OF LITTLE USE</th>
<th>OF GREAT USE</th>
</tr>
</thead>
</table>

Assuming that you could discover the answer to only one of these questions (B, C, or D), which would you ask in order to help you decide which company your parents rented the villa from?

Please circle your choice  B  C  D

306
Appendix 1c - Examples of the materials used in Experiment 3: Strong Information

First Condition

House Problem - Page 1

Your friend has just bought a new house. You can't remember whether it's on street X or street Y. You do remember that the house has a swimming pool and a garage. You have the following piece of information:

95% of houses on street X have swimming pools.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend now lives on street X and the other to complete certainty that your friend lives on street Y. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

1. What percentage of the houses on street Y would you expect to have a swimming pool?
2. What percentage of the houses on street X would you expect to have a garage?
3. What percentage of the houses on street Y would you expect to have a garage?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

House Problem - Page 2

Here is a second piece of information about the problem on the previous page:

70% of houses on street X have a garage.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.
Your friend is an engineer and works for a large construction company. It's either company A or company B, but you can't remember which. You do remember that he drives a company car and that he earns over £50,000 a year. You have the following piece of information:

95% of engineers working for company B earn over £50,000 a year.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend works for company A and the other to complete certainty that your friend works for company B. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

1. What percentage of engineers working for company A would you expect to earn over £50,000 a year?
2. What percentage of engineers working for company B would you expect to drive a company car?
3. What percentage of engineers working for company A would you expect to drive a company car?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Here is a second piece of information about the problem on the previous page:

70% of engineers working for company B drive a company car.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.
Hotel Problem - Page 1

Your friend is staying in a hotel in Paris. You can't remember whether it's hotel P or hotel Q. You do remember that the room costs £165 a night and that it has an en-suite bathroom. You have the following piece of information:

95% of the rooms in hotel P cost £165 a night.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend is staying in hotel P and the other to complete certainty that your friend is staying in hotel Q. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of the rooms in hotel Q would you expect to cost £165 a night?
(2)What percentage of the rooms in hotel P would you expect to have en-suite bathrooms?
(3) What percentage of the rooms in hotel Q would you expect to have en-suite bathrooms?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Hotel Problem - Page 2

Here is a second piece of information about the problem on the previous page:

70% of rooms in hotel P have an en-suite bathroom.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.
Appendix 1d - Examples of the materials used in Experiment 4: Low Association/Strong Information First Condition

House Problem - Page 1

Your friend has just bought a new house. You can't remember whether it's on street X or street Y. You do remember that the house has a swimming pool and that it was built sometime between 1945 and 1985. You have the following piece of information:

95% of houses on street X have swimming pools.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend now lives on street X and the other to complete certainty that your friend lives on street Y. Please mark the point on the line that corresponds to your belief about the alternatives.

Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of the houses on street Y would you expect to have a swimming pool?
(2) What percentage of the houses on street X would you expect to have been built sometime between 1945 and 1985?
(3) What percentage of the houses on street Y would you expect to have been built sometime between 1945 and 1985?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

House Problem - Page 2

Here is a second piece of information about the problem on the previous page:

70% of houses on street X were built sometime between 1945 and 1985.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of the houses on street Y would you expect to have a swimming pool?
(2) What percentage of the houses on street X would you expect to have been built sometime between 1945 and 1985?
(3) What percentage of the houses on street Y would you expect to have been built sometime between 1945 and 1985?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.
Engineer Problem - Page 1

Your friend is a civil engineer and works for a large construction company. It's either company A or company B, but you can't remember which. You do remember that he works in general construction and that he earns over £60,000 a year. You have the following piece of information:

*95% of civil engineers working for company A earn over £60,000 a year.*

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend works for company A and the other to complete certainty that your friend works for company B. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

![Scale Image]

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of civil engineers working for company B would you expect to earn over £60,000 a year?

(2) What percentage of civil engineers working for company A would you expect to work in general construction?

(3) What percentage of civil engineers working for company B would you expect to work in general construction?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Engineer Problem - Page 2

Here is a second piece of information about the problem on the previous page:

*70% of civil engineers working for company A work in general construction.*

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

![Scale Image]
Car Problem - Page 1

Your sister bought a new car in 1988. You can't remember whether it's a model X or a model Y. You do remember that it has four doors and a top speed of over 165 miles per hour. You have the following piece of information:

95% of model Xs bought in 1988 have a top speed of over 165 mph.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your sister's car is a model X and the other to complete certainty that her car is a model Y. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of model Ys bought in 1988 would you expect to have a top speed of over 165 mph?
(2) What percentage of model Xs bought in 1988 would you expect to have four doors?
(3) What percentage of model Ys bought in 1988 would you expect to have four doors?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Car Problem - Page 2

Here is a second piece of information about the problem on the previous page:

70% of model Xs bought in 1988 have four doors.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.
Appendix 1e - Examples of stimulus materials from Experiment 6: Rare Features/Strong Information First Condition

House Problem - Page 1

Your friend has just bought a new house. You can't remember whether its' on street X or street Y. You do remember that the house has a swimming pool and that it's worth approximately £250,000. You have the following piece of information:

95% of houses on street X have swimming pools.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend now lives on street X and the other to complete certainty that your friend lives on street Y. Please mark the point on the line that corresponds to your belief about the alternatives.

Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Certain that my friend lives on Street Y  Certain that my friend lives on Street X

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of the houses on street Y would you expect to have a swimming pool?
(2) What percentage of the houses on street X would you expect to be worth approximately £250,000?
(3) What percentage of the houses on street Y would you expect to be worth approximately £250,000?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

House Problem - Page 2

Here is a second piece of information about the problem on the previous page:

70% of houses on street X are worth approximately £250,000.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

Certain that my friend lives on Street Y  Certain that my friend lives on Street X

313
Your friend is a civil engineer and works for a large construction company. It's either company A or company B, but you can't remember which. You do remember that he travels regularly to supervise projects abroad and that he earns over £60,000 a year. You have the following piece of information:

95% of civil engineers working for company A earn over £60,000 a year.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend works for company A and the other to complete certainty that your friend works for company B. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Certain that my friend works for company Y

Certain that my friend works for company X

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of civil engineers working for company B would you expect to earn over £60,000 a year?
(2) What percentage of civil engineers working for company A would you expect to travel regularly to supervise projects abroad?
(3) What percentage of civil engineers working for company B would you expect to travel regularly to supervise projects abroad?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Here is a second piece of information about the problem on the previous page:

70% of civil engineers working for company A travel regularly to supervise projects abroad.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

Certain that my friend works for company Y

Certain that my friend works for company X
Your sister bought a new car in 1988. You can’t remember whether its’ a model X or a model Y. You do remember that it has leather upholstery and a top speed of over 165 miles per hour. You have the following piece of information:

95% of model Xs bought in 1988 have a top speed of over 165 mph.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your sister's car is a model X and the other to complete certainty that her car is a model Y. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of model Ys bought in 1988 would you expect to have a top speed of over 165 mph?
(2) What percentage of model Xs bought in 1988 would you expect to have leather upholstery?
(3) What percentage of model Ys bought in 1988 would you expect to have leather upholstery?

Here is a second piece of information about the problem on the previous page:

70% of model Xs bought in 1988 have leather upholstery.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.
Appendix 1f - Example materials from Experiment 7: Strong Information First Condition

Adjective/Passive Verb Problem - Page 1
You are currently studying a Zogese symbol. The symbol definitely represents either an adjective or a passive verb, and your task is to decide which of these two parts of speech the symbol actually represents. The symbol itself is triangular and there is a diagonal line running through the triangle. You have the following piece of information:

95% of adjectives in Zogese have a triangular element.

Now look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that the symbol you are studying represents a passive verb and the other to complete certainty that it represents an adjective. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of passive verbs in Zogese would you expect to have a triangular element?
(2) What percentage of adjectives in Zogese would you expect to have a diagonal line running through the main part of the symbol?
(3) What percentage of passive verbs in Zogese would you expect to have a diagonal line running through the main part of the symbol?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Adjective/Passive Verb Problem - Page 2
Here is a second piece of information about the problem on the previous page:

70% of adjectives in Zogese have a diagonal line running through the main part of the symbol.

Once again there is a scale below these instructions that is designed to measure your confidence in each of the decision alternatives in the light of the information available to you. Once again either end of the scale corresponds to complete certainty that one or other of the decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.
Proper Noun/Preposition Problem - Page 1

You are currently studying a Zogese symbol. The symbol definitely represents either a proper noun or a preposition, and your task is to decide which of these two parts of speech the symbol actually represents. The symbol itself is circular and the circular part of the symbol is followed by a series of three dots. You have the following piece of information:

95% of proper nouns in Zogese have a circular element.

Now look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that the symbol you are studying represents a proper noun and the other to complete certainty that it represents a preposition. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

```
certain that the symbol represents a preposition  certain that the symbol represents a proper noun
```

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of prepositions in Zogese would you expect to have a circular element?
(2) What percentage of proper nouns in Zogese would you expect to have a series of three dots after the main part of the symbol?
(3) What percentage of prepositions in Zogese would you expect to have a series of three dots after the main part of the symbol?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Proper Noun/Preposition Problem - Page 2

Here is a second piece of information about the problem on the previous page:

70% of proper nouns in Zogese have a series of three dots after the main part of the symbol.

Once again there is a scale below these instructions that is designed to measure your confidence in each of the decision alternatives in the light of the information available to you. Once again either end of the scale corresponds to complete certainty that one or other of the decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

```
certain that the symbol represents a preposition  certain that the symbol represents a proper noun
```

317
You are currently studying a Zogese symbol. The symbol definitely represents either an active verb or a pronoun, and your task is to decide which it actually is. The symbol itself is rectangular and there is a line over the main part of the symbol. You have the following piece of information:

95% of pronouns in Zogese have a rectangular element.

Now look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that the symbol you are studying represents a pronoun and the other to complete certainty that it represents an active verb. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

(1) What percentage of active verbs in Zogese would you expect to have a rectangular element?
(2) What percentage of pronouns in Zogese would you expect to have a line over the main part of the symbol?
(3) What percentage of active verbs in Zogese would you expect to have a line over the main part of the symbol?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Here is a second piece of information about the problem on the previous page:

70% of pronouns in Zogese have a line over the main part of the symbol.

Once again there is a scale below these instructions that is designed to measure your confidence in each of the decision alternatives in the light of the information available to you. Once again either end of the scale corresponds to complete certainty that one or other of the decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.
House Problem - Page 1

Your friend has just bought a new house. You can't remember whether it's on street X or street Y. You do remember that the house has a swimming pool and a garage. You have the following piece of information:

95% of houses on street X have a swimming pool.

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend now lives on street X and the other to complete certainty that your friend lives on street Y. Please mark the point on the line that corresponds to your belief about the alternatives.

Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

Once you have placed your mark on the line please answer the following questions:

1. What percentage of the houses on street Y would you expect to have a swimming pool?
2. What percentage of the houses on street X would you expect to have a garage?
3. What percentage of the houses on street Y would you expect to have a garage?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

House Problem - Page 2

Here is a second piece of information about the problem on the previous page:

25% of houses on street Y have a swimming pool.

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.
Engineer Problem - Page 1

Your friend is a civil engineer and works for a large construction company. It's either company A or company B, but you can't remember which. You do remember that he drives a company car and that he earns over £60,000 a year. You have the following piece of information:

**95% of civil engineers working for company A earn over £60,000 a year.**

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your friend works for company A and the other to complete certainty that your friend works for company B. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

[Scale image]

Certain that my friend works for company Y  Certain that my friend works for company X

Once you have placed your mark on the line please answer the following questions:

1. What percentage of civil engineers working for company B would you expect to earn over £60,000 a year?
2. What percentage of civil engineers working for company A would you expect to drive a company car?
3. What percentage of civil engineers working for company B would you expect to drive a company car?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

Engineer Problem - Page 2

Here is a second piece of information about the problem on the previous page:

**25% of civil engineers working for company B earn over £60,000 a year.**

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

[Scale image]

Certain that my friend works for company Y  Certain that my friend works for company X
Your sister bought a new car in 1988. You can't remember whether its' a model X or a model Y. You do remember that it has a radio and a top speed of over 165 miles per hour. You have the following piece of information:

**95% of model Xs bought in 1988 have a top speed of over 165 mph.**

Now please look at the scale below which is designed to measure your confidence in each of the decision alternatives in the light of the information you have available to you. One end of the scale corresponds to complete certainty that your sister’s car is a model X and the other to complete certainty that her car is a model Y. Please mark the point on the line that corresponds to your belief about the alternatives. Remember that the greater your confidence in an alternative the closer your mark should be to the end of the scale that corresponds to complete certainty in that alternative.

<table>
<thead>
<tr>
<th>Certain that my sister's car is a model Y</th>
<th>Certain that my sister's car is a model X</th>
</tr>
</thead>
</table>

Once you have placed your mark on the line please answer the following questions:

1. What percentage of model Ys bought in 1988 would you expect to have a top speed of over 165 mph?
2. What percentage of model Xs bought in 1988 would you expect to have a radio?
3. What percentage of model Ys bought in 1988 would you expect to have a radio?

When you have finished answering the questions separate this page from the one behind it by gently pulling the sheets apart at the bottom.

**Car Problem - Page 2**

Here is a second piece of information about the problem on the previous page:

**25% of model Ys bought in 1988 have a top speed of over 165 mph.**

Once again there is a scale below these instructions designed to measure your confidence in each of the decision alternatives in light of the information you have available to you. Once again either end of the scale corresponds to complete certainty that one or the other of decision alternatives is true. Bearing the new piece of information in mind, please mark the point on the line that corresponds to your belief about the alternatives.

<table>
<thead>
<tr>
<th>Certain that my sister's car is a model Y</th>
<th>Certain that my sister's car is a model X</th>
</tr>
</thead>
</table>
APPENDIX 2 - Summary tables, tables of means and frequency tables from Experiments 1 - 4

Table of Contents

Appendix 2a - Selection frequencies for Experiment 1 ................................................. 323

Appendix 2b - Anova tables and tables of means from the analyses of subjects' confidence ratings in Experiment 1 ........................................................... 324

Appendix 2c - Selection frequencies from Experiment 2 ................................................. 326

Appendix 2d - Anova tables from the analyses of subjects' evidence selections and usefulness ratings in Experiment 2 ........................................................... 327

Appendix 2e - Table of means and standard deviations from the analysis of subjects' usefulness ratings in Experiment 2 ........................................................... 328

Appendix 2f - Anova tables from the analyses of subjects' confidence ratings and expectations in Experiment 3 ........................................................... 329

Appendix 2g - Anova tables from the analyses of subjects' confidence ratings and expectations in Experiment 4 ........................................................... 330
# Appendix 2a - Selection frequencies for Experiment 1

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Rarity</th>
<th>Engineer</th>
<th>University</th>
<th>Hotel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>65%</td>
<td>Rare</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Comm</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>95%</td>
<td>Rare</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Comm</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>100%</td>
<td>Rare</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Comm</td>
<td>1</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
### Appendix 2b - Anova tables and tables of means from the analyses of subjects' confidence ratings in Experiment 1

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4834.898</td>
<td>51</td>
<td>469.8523</td>
<td>10.29025</td>
<td>.0023122</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>634.689</td>
<td>51</td>
<td>469.8523</td>
<td>1.35803</td>
<td>.2505438</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>427.299</td>
<td>51</td>
<td>469.8523</td>
<td>.90943</td>
<td>.3447635</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>85.422</td>
<td>51</td>
<td>469.8523</td>
<td>.18181</td>
<td>.6716197</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>137.521</td>
<td>51</td>
<td>469.8523</td>
<td>.29269</td>
<td>.5908573</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>576.022</td>
<td>51</td>
<td>469.8523</td>
<td>1.22596</td>
<td>.2733883</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>718.223</td>
<td>51</td>
<td>469.8523</td>
<td>1.52862</td>
<td>.2219843</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>3332.023</td>
<td>51</td>
<td>469.8523</td>
<td>7.09164</td>
<td>.0103367</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>1781.153</td>
<td>51</td>
<td>469.8523</td>
<td>3.79088</td>
<td>.0570496</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>1.712</td>
<td>51</td>
<td>469.8523</td>
<td>.00364</td>
<td>.9520995</td>
</tr>
<tr>
<td>123</td>
<td>1</td>
<td>378.772</td>
<td>51</td>
<td>469.8523</td>
<td>.80615</td>
<td>.3734797</td>
</tr>
<tr>
<td>124</td>
<td>1</td>
<td>82.689</td>
<td>51</td>
<td>469.8523</td>
<td>.17599</td>
<td>.6766049</td>
</tr>
<tr>
<td>134</td>
<td>1</td>
<td>387.200</td>
<td>51</td>
<td>469.8523</td>
<td>.82409</td>
<td>.3682571</td>
</tr>
<tr>
<td>234</td>
<td>1</td>
<td>2158.580</td>
<td>51</td>
<td>469.8523</td>
<td>4.59417</td>
<td>.0368727</td>
</tr>
<tr>
<td>1234</td>
<td>1</td>
<td>157.631</td>
<td>51</td>
<td>469.8523</td>
<td>.33549</td>
<td>.5649939</td>
</tr>
</tbody>
</table>

2b(i): The effect of Likelihood (=1), Rarity (=2), Feedback (=3) and Choice on the confidence ratings of subjects in the 65% and 95% Likelihood conditions of Experiment 1.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>2</td>
<td>558.3305</td>
<td>64</td>
<td>647.9248</td>
<td>.8617211</td>
<td>.4272770</td>
</tr>
</tbody>
</table>

2b(ii): The effect of problem content on subjects' confidence ratings in Experiment 1.
### Positive Feedback

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Choice</th>
<th>Rare Feature</th>
<th>Comm Feature</th>
<th>Rare Feature</th>
<th>Comm Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>65%</td>
<td>Diagnostic</td>
<td>59 30 N = 5</td>
<td>43 23 N = 5</td>
<td>36 10 N = 4</td>
<td>33 12 N = 6</td>
</tr>
<tr>
<td></td>
<td>Pseudodiag</td>
<td>49 24 N = 3</td>
<td>76 17 N = 3</td>
<td>54 15 N = 4</td>
<td>60 41 N = 3</td>
</tr>
<tr>
<td>95%</td>
<td>Diagnostic</td>
<td>73 8 N = 4</td>
<td>71 22 N = 3</td>
<td>76 11 N = 4</td>
<td>80 6 N = 6</td>
</tr>
<tr>
<td></td>
<td>Pseudodiag</td>
<td>42 32 N = 4</td>
<td>87 14 N = 5</td>
<td>65 28 N = 5</td>
<td>56 37 N = 3</td>
</tr>
</tbody>
</table>

2b(iii): Means (in bold), standard deviations and subject numbers from the analysis of the effects of Likelihood, Rarity, Feedback and Choice on subjects' confidence ratings in the 65% and 95% Likelihood conditions of Experiment 1.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>301.014</td>
<td>23</td>
<td>843.1418</td>
<td>.357014</td>
<td>.556068</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4829.014</td>
<td>23</td>
<td>843.1418</td>
<td>5.727404</td>
<td>.0252548</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1163.905</td>
<td>23</td>
<td>843.1418</td>
<td>1.380438</td>
<td>.2520474</td>
</tr>
</tbody>
</table>

2b(iv): The effect of Rarity (=1) and Feedback (=2) on the confidence ratings of subjects in the 100% likelihood condition who choose the pseudodiagnostic evidence in Experiment 1.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>2</td>
<td>69.58393</td>
<td>24</td>
<td>1047.473</td>
<td>.0654303</td>
<td>.9358996</td>
</tr>
</tbody>
</table>

2b(v): The effect problem content on the confidence ratings of subjects in the 100% likelihood condition who choose the pseudodiagnostic evidence in Experiment 1.

### Negative Feedback

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Choice</th>
<th>Rare Feature</th>
<th>Comm Feature</th>
<th>Rare Feature</th>
<th>Comm Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>65%</td>
<td>Diagnostic</td>
<td>59 30 N = 5</td>
<td>43 23 N = 5</td>
<td>36 10 N = 4</td>
<td>33 12 N = 6</td>
</tr>
<tr>
<td></td>
<td>Pseudodiag</td>
<td>49 24 N = 3</td>
<td>76 17 N = 3</td>
<td>54 15 N = 4</td>
<td>60 41 N = 3</td>
</tr>
<tr>
<td>95%</td>
<td>Diagnostic</td>
<td>73 8 N = 4</td>
<td>71 22 N = 3</td>
<td>76 11 N = 4</td>
<td>80 6 N = 6</td>
</tr>
<tr>
<td></td>
<td>Pseudodiag</td>
<td>42 32 N = 4</td>
<td>87 14 N = 5</td>
<td>65 28 N = 5</td>
<td>56 37 N = 3</td>
</tr>
</tbody>
</table>

2b(vi): Means (in bold), standard deviations and subject numbers from the analysis of the effects of Rarity and Feedback on the confidence ratings of subjects in the 100% Likelihood condition who selected pseudodiagnostic evidence in Experiment 1.
## Appendix 2c - Selection frequencies from Experiment 2

<table>
<thead>
<tr>
<th>Content</th>
<th>Choice</th>
<th>Common Features</th>
<th>Rare Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hint</td>
<td>No Hint</td>
</tr>
<tr>
<td>House B</td>
<td>5</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Engineer B</td>
<td>6</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Car B</td>
<td>5</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Villa B</td>
<td>8</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>
Appendix 2d - Anova tables from the analyses of subjects' evidence selections and usefulness ratings in Experiment 2

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2.263973</td>
<td>90</td>
<td>2.258754</td>
<td>1.002311</td>
<td>.3194362</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9.730639</td>
<td>90</td>
<td>2.258754</td>
<td>4.307967</td>
<td>.0407876</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4.530640</td>
<td>90</td>
<td>2.258754</td>
<td>2.005814</td>
<td>.1601490</td>
</tr>
</tbody>
</table>

2d(i): The effect of Hint (=1) and Rarity (=2) on subjects' selection of item C in Experiment 2.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2501.44</td>
<td>90</td>
<td>2291.684</td>
<td>1.09153</td>
<td>.2989305</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3453.96</td>
<td>90</td>
<td>2291.684</td>
<td>1.50717</td>
<td>.2227727</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1125.62</td>
<td>270</td>
<td>415.721</td>
<td>2.70762</td>
<td>.0456445</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>19693.05</td>
<td>180</td>
<td>1094.553</td>
<td>17.99187</td>
<td>.0000001</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>416.76</td>
<td>90</td>
<td>2291.684</td>
<td>.18186</td>
<td>.6708001</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>444.98</td>
<td>270</td>
<td>415.721</td>
<td>1.07038</td>
<td>.3620817</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>391.19</td>
<td>270</td>
<td>415.721</td>
<td>.94100</td>
<td>.4212760</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2952.17</td>
<td>180</td>
<td>1094.553</td>
<td>2.69714</td>
<td>.0701219</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>1018.51</td>
<td>180</td>
<td>1094.553</td>
<td>.93053</td>
<td>.3962351</td>
</tr>
<tr>
<td>34</td>
<td>6</td>
<td>918.16</td>
<td>540</td>
<td>359.861</td>
<td>2.55142</td>
<td>.0191122</td>
</tr>
<tr>
<td>123</td>
<td>3</td>
<td>403.54</td>
<td>270</td>
<td>415.721</td>
<td>.97070</td>
<td>.4070068</td>
</tr>
<tr>
<td>124</td>
<td>2</td>
<td>2722.54</td>
<td>180</td>
<td>1094.553</td>
<td>2.48735</td>
<td>.0859835</td>
</tr>
<tr>
<td>134</td>
<td>6</td>
<td>378.48</td>
<td>540</td>
<td>359.861</td>
<td>1.05173</td>
<td>.3907560</td>
</tr>
<tr>
<td>234</td>
<td>6</td>
<td>513.03</td>
<td>540</td>
<td>359.861</td>
<td>1.42565</td>
<td>.2025427</td>
</tr>
<tr>
<td>1234</td>
<td>6</td>
<td>704.99</td>
<td>540</td>
<td>359.861</td>
<td>1.95906</td>
<td>.0697189</td>
</tr>
</tbody>
</table>

2d(ii): The effect of Hint (=1), Rarity (=2), problem content (=3) and Item Rated (=4) on subjects' usefulness ratings in Experiment 2.
Appendix 2e - Table of means and standard deviations from the analysis of subjects' usefulness ratings in Experiment 2

<table>
<thead>
<tr>
<th>Content</th>
<th>Item Rated</th>
<th>Common Features</th>
<th></th>
<th>Rare Features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hint</td>
<td>No Hint</td>
<td>Hint</td>
<td>No Hint</td>
</tr>
<tr>
<td>House</td>
<td>B</td>
<td>64 27</td>
<td>67 26</td>
<td>67 25</td>
<td>70 27</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>82 19</td>
<td>75 24</td>
<td>69 24</td>
<td>66 29</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>59 27</td>
<td>72 20</td>
<td>59 26</td>
<td>54 26</td>
</tr>
<tr>
<td>Engineer</td>
<td>B</td>
<td>59 28</td>
<td>70 24</td>
<td>58 29</td>
<td>73 23</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>79 19</td>
<td>73 18</td>
<td>61 32</td>
<td>67 31</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>51 27</td>
<td>61 25</td>
<td>51 29</td>
<td>51 26</td>
</tr>
<tr>
<td>Car</td>
<td>B</td>
<td>54 28</td>
<td>66 26</td>
<td>58 30</td>
<td>56 33</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>83 15</td>
<td>67 24</td>
<td>67 25</td>
<td>78 16</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>53 26</td>
<td>60 26</td>
<td>53 28</td>
<td>67 22</td>
</tr>
<tr>
<td>Villa</td>
<td>B</td>
<td>64 27</td>
<td>69 23</td>
<td>56 31</td>
<td>72 21</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>76 22</td>
<td>64 27</td>
<td>67 28</td>
<td>68 23</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>55 26</td>
<td>58 27</td>
<td>56 26</td>
<td>51 26</td>
</tr>
</tbody>
</table>

(Means in bold)
Appendix 2f - Anova tables from the analyses of subjects' confidence ratings and expectations in Experiment 3

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1485.520</td>
<td>56</td>
<td>672.9253</td>
<td>2.20756</td>
<td>.1429445</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2061.520</td>
<td>56</td>
<td>212.0218</td>
<td>9.72315</td>
<td>.0028722</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>187.940</td>
<td>112</td>
<td>295.1402</td>
<td>.63678</td>
<td>.5308968</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4454.095</td>
<td>56</td>
<td>212.0218</td>
<td>21.00772</td>
<td>.0000261</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>44.710</td>
<td>112</td>
<td>295.1402</td>
<td>.15149</td>
<td>.8596053</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>276.003</td>
<td>112</td>
<td>115.1160</td>
<td>2.39761</td>
<td>.0955887</td>
</tr>
<tr>
<td>123</td>
<td>2</td>
<td>146.336</td>
<td>112</td>
<td>115.1160</td>
<td>1.27121</td>
<td>.2845082</td>
</tr>
</tbody>
</table>

2f(i): The effect of Order (=1), Before/After (=2) and problem content on subjects' confidence ratings in Experiment 3.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>47251.31</td>
<td>33</td>
<td>805.2808</td>
<td>58.67682</td>
<td>.0000000</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>858.02</td>
<td>66</td>
<td>440.6500</td>
<td>1.94716</td>
<td>.1507881</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2759.16</td>
<td>66</td>
<td>440.6500</td>
<td>6.26157</td>
<td>.0032361</td>
</tr>
</tbody>
</table>

2f(ii): The effect of Order (=1) and problem content on subjects' expectations in Experiment 3.
Appendix 2g - Anova tables from the analyses of subjects' confidence ratings and expectations in Experiment 4

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>145.836</td>
<td>68</td>
<td>1239.351</td>
<td>.11767</td>
<td>.7326326</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.058</td>
<td>68</td>
<td>1239.351</td>
<td>.00005</td>
<td>.9945679</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5090.947</td>
<td>68</td>
<td>231.234</td>
<td>22.01647</td>
<td>.0000135</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>102.502</td>
<td>136</td>
<td>121.566</td>
<td>.84318</td>
<td>.4325772</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>479.225</td>
<td>68</td>
<td>1239.351</td>
<td>.38667</td>
<td>.5361324</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>167.502</td>
<td>68</td>
<td>231.234</td>
<td>.72439</td>
<td>.3976963</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>2675.058</td>
<td>68</td>
<td>231.234</td>
<td>11.56864</td>
<td>.0011270</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>218.155</td>
<td>136</td>
<td>121.566</td>
<td>1.79453</td>
<td>.1701180</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>155.530</td>
<td>136</td>
<td>121.566</td>
<td>1.27938</td>
<td>.2815355</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>13.766</td>
<td>136</td>
<td>88.493</td>
<td>.15556</td>
<td>.8560854</td>
</tr>
<tr>
<td>123</td>
<td>1</td>
<td>1048.447</td>
<td>68</td>
<td>231.234</td>
<td>4.53415</td>
<td>.0368485</td>
</tr>
<tr>
<td>124</td>
<td>2</td>
<td>12.461</td>
<td>136</td>
<td>121.566</td>
<td>.10250</td>
<td>.9026471</td>
</tr>
<tr>
<td>134</td>
<td>2</td>
<td>75.502</td>
<td>136</td>
<td>88.493</td>
<td>.85320</td>
<td>.4283162</td>
</tr>
<tr>
<td>234</td>
<td>2</td>
<td>25.488</td>
<td>136</td>
<td>88.493</td>
<td>.28803</td>
<td>.7501969</td>
</tr>
<tr>
<td>1234</td>
<td>2</td>
<td>74.558</td>
<td>136</td>
<td>88.493</td>
<td>.84253</td>
<td>.4328550</td>
</tr>
</tbody>
</table>

2g(i): The effects of Association (=1), Order (= 2), Before/After (=3) and problem content (=4) on subjects' confidence ratings in Experiment 4.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>7566.23</td>
<td>52</td>
<td>1449.686</td>
<td>5.21922</td>
<td>.0264495</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>53342.57</td>
<td>52</td>
<td>1449.686</td>
<td>36.79595</td>
<td>.0000002</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>624.46</td>
<td>104</td>
<td>387.796</td>
<td>1.61029</td>
<td>.2047711</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1831.57</td>
<td>52</td>
<td>1449.686</td>
<td>1.26342</td>
<td>.2661665</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>540.77</td>
<td>104</td>
<td>387.796</td>
<td>1.39447</td>
<td>.2525614</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>1294.95</td>
<td>104</td>
<td>387.796</td>
<td>3.33925</td>
<td>.0393046</td>
</tr>
<tr>
<td>123</td>
<td>2</td>
<td>875.58</td>
<td>104</td>
<td>387.796</td>
<td>2.25783</td>
<td>.1096801</td>
</tr>
</tbody>
</table>

2g(ii): The effects of Association (=1), Order (=2) and problem content (=3) on subjects' expectations in Experiment 4.
APPENDIX 3 - Summary tables from the analyses of subjects' confidence ratings and expectations in Experiments 5 - 8

Table of Contents

Appendix 3a - Anova tables from the analyses of subjects' confidence ratings and expectations in Experiment 5 ................................................................. 332

Appendix 3b - Anova tables for the analyses of subjects' confidence ratings and expectations in Experiment 6 ................................................................. 333

Appendix 3c - Anova tables for the analyses of subjects' confidence ratings and expectations in Experiment 7 ................................................................. 334

Appendix 3d - Anova table from the analysis of subjects' confidence ratings in Experiment 8 ................................................................. 335
Appendix 3a - Anova tables from the analyses of subjects' confidence ratings and expectations in Experiment 5.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2710.958</td>
<td>67</td>
<td>1022.009</td>
<td>2.65258</td>
<td>.1080757</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>116.397</td>
<td>67</td>
<td>1022.009</td>
<td>.11389</td>
<td>.7368131</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5901.179</td>
<td>67</td>
<td>157.445</td>
<td>37.48096</td>
<td>.0000001</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>193.125</td>
<td>134</td>
<td>186.956</td>
<td>1.03300</td>
<td>.3587555</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1526.011</td>
<td>67</td>
<td>1022.009</td>
<td>1.49315</td>
<td>.2260105</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>560.430</td>
<td>67</td>
<td>157.445</td>
<td>3.55953</td>
<td>.0635384</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>1344.309</td>
<td>67</td>
<td>157.445</td>
<td>8.53830</td>
<td>.0000001</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>15.238</td>
<td>134</td>
<td>186.956</td>
<td>.08151</td>
<td>.9217724</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>179.570</td>
<td>134</td>
<td>186.956</td>
<td>.96049</td>
<td>.3853229</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>122.858</td>
<td>134</td>
<td>68.800</td>
<td>1.78571</td>
<td>.1716448</td>
</tr>
<tr>
<td>123</td>
<td>1</td>
<td>508.990</td>
<td>67</td>
<td>157.445</td>
<td>3.23282</td>
<td>.0766817</td>
</tr>
<tr>
<td>124</td>
<td>2</td>
<td>26.454</td>
<td>134</td>
<td>186.956</td>
<td>.14150</td>
<td>.8681865</td>
</tr>
<tr>
<td>134</td>
<td>2</td>
<td>70.823</td>
<td>134</td>
<td>68.800</td>
<td>1.02940</td>
<td>.3600276</td>
</tr>
<tr>
<td>234</td>
<td>2</td>
<td>88.268</td>
<td>134</td>
<td>68.800</td>
<td>1.28296</td>
<td>.2805980</td>
</tr>
<tr>
<td>1234</td>
<td>2</td>
<td>55.174</td>
<td>134</td>
<td>68.800</td>
<td>.80195</td>
<td>.4505948</td>
</tr>
</tbody>
</table>

3a(i): The effects of Likelihood (=1), Order (=2), Before/After (=3) and problem content (=4) on subjects' confidence ratings in Experiment 5.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3295.20</td>
<td>54</td>
<td>1010.879</td>
<td>3.25974</td>
<td>.0765760</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>50877.82</td>
<td>54</td>
<td>1010.879</td>
<td>50.33029</td>
<td>.0000000</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>998.16</td>
<td>108</td>
<td>468.231</td>
<td>2.13178</td>
<td>.1235914</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>89.36</td>
<td>54</td>
<td>1010.879</td>
<td>.08840</td>
<td>.7673629</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1118.45</td>
<td>108</td>
<td>468.231</td>
<td>2.38866</td>
<td>.0965844</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>3337.48</td>
<td>108</td>
<td>468.231</td>
<td>7.12785</td>
<td>.0012369</td>
</tr>
<tr>
<td>123</td>
<td>2</td>
<td>253.41</td>
<td>108</td>
<td>468.231</td>
<td>.54120</td>
<td>.5836208</td>
</tr>
</tbody>
</table>

3a(ii): The effects of Likelihood (=1), Order (=2) and problem content on subjects' expectations in Experiment 5.
### Appendix 3b - Anova tables for the analyses of subjects' confidence ratings and expectations in Experiment 6.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>665.18</td>
<td>66</td>
<td>844.9006</td>
<td>.78728</td>
<td>.3781447</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>15.96</td>
<td>66</td>
<td>844.9006</td>
<td>.01888</td>
<td>.8911160</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>12299.24</td>
<td>66</td>
<td>172.9042</td>
<td>71.13326</td>
<td>.000000</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>10.95</td>
<td>132</td>
<td>176.6656</td>
<td>.06196</td>
<td>.9399454</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>968.03</td>
<td>66</td>
<td>844.9006</td>
<td>1.14574</td>
<td>.2883422</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1205.67</td>
<td>66</td>
<td>172.9042</td>
<td>6.97305</td>
<td>.0103181</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>30.94</td>
<td>66</td>
<td>172.9042</td>
<td>.17893</td>
<td>.6736727</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>98.76</td>
<td>132</td>
<td>176.6656</td>
<td>.55904</td>
<td>.5731082</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>168.97</td>
<td>132</td>
<td>176.6656</td>
<td>.95642</td>
<td>.3869144</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>100.33</td>
<td>132</td>
<td>88.2156</td>
<td>1.13733</td>
<td>.3237953</td>
</tr>
<tr>
<td>123</td>
<td>1</td>
<td>241.57</td>
<td>66</td>
<td>172.9042</td>
<td>1.39716</td>
<td>.2414410</td>
</tr>
<tr>
<td>124</td>
<td>2</td>
<td>.56</td>
<td>132</td>
<td>176.6656</td>
<td>.00318</td>
<td>.9968225</td>
</tr>
<tr>
<td>134</td>
<td>2</td>
<td>23.80</td>
<td>132</td>
<td>88.2156</td>
<td>.26980</td>
<td>.7639527</td>
</tr>
<tr>
<td>234</td>
<td>2</td>
<td>162.74</td>
<td>132</td>
<td>88.2156</td>
<td>1.84480</td>
<td>.1621091</td>
</tr>
<tr>
<td>1234</td>
<td>2</td>
<td>184.22</td>
<td>132</td>
<td>88.2156</td>
<td>2.08829</td>
<td>.1279737</td>
</tr>
</tbody>
</table>

**3b(i):** The effects of Order (=1), Rarity (=2), Before/After (=3) and problem content (=4) on subjects' confidence ratings in Experiment 6.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>407.748</td>
<td>55</td>
<td>516.0072</td>
<td>.790198</td>
<td>.3779116</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2188.466</td>
<td>55</td>
<td>516.0072</td>
<td>4.241155</td>
<td>.0442017</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3932.718</td>
<td>110</td>
<td>434.9565</td>
<td>9.041636</td>
<td>.0002315</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>722.661</td>
<td>55</td>
<td>516.0072</td>
<td>1.400487</td>
<td>.2417320</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1256.651</td>
<td>110</td>
<td>434.9565</td>
<td>2.889142</td>
<td>.0598559</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>295.105</td>
<td>110</td>
<td>434.9565</td>
<td>.678470</td>
<td>.5095030</td>
</tr>
<tr>
<td>123</td>
<td>2</td>
<td>77.009</td>
<td>110</td>
<td>434.9565</td>
<td>.177049</td>
<td>.8379769</td>
</tr>
</tbody>
</table>

**3b(ii):** The effects of Order (=1), Rarity (=2) and problem content (=3) on subjects' expectations in Experiment 6.
### Appendix 3c - Anova tables for the analyses of subjects' confidence ratings and expectations in Experiment 7.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>426.96</td>
<td>33</td>
<td>1154.009</td>
<td>.36998</td>
<td>.5471779</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>10275.20</td>
<td>33</td>
<td>118.286</td>
<td>86.86724</td>
<td>.0000000</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>37.31</td>
<td>66</td>
<td>73.988</td>
<td>.50433</td>
<td>.6062158</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1793.68</td>
<td>33</td>
<td>118.286</td>
<td>15.16387</td>
<td>.0004540</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>98.43</td>
<td>66</td>
<td>73.988</td>
<td>1.33034</td>
<td>.2713834</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>28.10</td>
<td>66</td>
<td>49.700</td>
<td>.56547</td>
<td>.5708208</td>
</tr>
<tr>
<td>123</td>
<td>2</td>
<td>90.09</td>
<td>66</td>
<td>49.700</td>
<td>1.81275</td>
<td>.1712357</td>
</tr>
</tbody>
</table>

3c(i): The effects of Order (=1), Before/After (=2) and problem content on subjects' confidence ratings in Experiment 7.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>126.2976</td>
<td>26</td>
<td>736.8690</td>
<td>.171398</td>
<td>.6822681</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>465.2262</td>
<td>52</td>
<td>443.4240</td>
<td>1.049169</td>
<td>.3575242</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>402.0833</td>
<td>52</td>
<td>443.4240</td>
<td>.906769</td>
<td>.4101158</td>
</tr>
</tbody>
</table>

3c(ii): The effects of Order (=1) and problem content (=2) on subjects' expectations in Experiment 7.
## Appendix 3d - Anova table from the analysis of subjects' confidence ratings in Experiment 8

<table>
<thead>
<tr>
<th>Effect</th>
<th>df Effect</th>
<th>MS Effect</th>
<th>df Error</th>
<th>MS Error</th>
<th>F ratio</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8199.12</td>
<td>135</td>
<td>484.7120</td>
<td>16.9154</td>
<td>.0000675</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1927.88</td>
<td>135</td>
<td>484.7120</td>
<td>3.9774</td>
<td>.0481295</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6138.14</td>
<td>135</td>
<td>484.7120</td>
<td>12.6635</td>
<td>.0005152</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>65997.64</td>
<td>135</td>
<td>364.1313</td>
<td>181.2468</td>
<td>.0000000</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>443.64</td>
<td>270</td>
<td>121.7271</td>
<td>3.6446</td>
<td>.0274268</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1612.44</td>
<td>135</td>
<td>484.7120</td>
<td>3.3266</td>
<td>.0703799</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>921.07</td>
<td>135</td>
<td>484.7120</td>
<td>1.9002</td>
<td>.1703311</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>42.07</td>
<td>135</td>
<td>484.7120</td>
<td>.0868</td>
<td>.7687454</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>27260.94</td>
<td>135</td>
<td>364.1313</td>
<td>74.8657</td>
<td>.0000000</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>0.00</td>
<td>135</td>
<td>364.1313</td>
<td>0.0000</td>
<td>.9985853</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>341.27</td>
<td>135</td>
<td>364.1313</td>
<td>.9372</td>
<td>.3347261</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>15.88</td>
<td>270</td>
<td>121.7271</td>
<td>.1305</td>
<td>.8777403</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>51.43</td>
<td>270</td>
<td>121.7271</td>
<td>.4225</td>
<td>.6558170</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>120.12</td>
<td>270</td>
<td>121.7271</td>
<td>.9868</td>
<td>.3740992</td>
</tr>
<tr>
<td>45</td>
<td>2</td>
<td>10.33</td>
<td>270</td>
<td>92.5111</td>
<td>.1116</td>
<td>.8944135</td>
</tr>
<tr>
<td>123</td>
<td>1</td>
<td>962.68</td>
<td>135</td>
<td>484.7120</td>
<td>1.9861</td>
<td>.1610502</td>
</tr>
<tr>
<td>124</td>
<td>1</td>
<td>1527.44</td>
<td>135</td>
<td>364.1313</td>
<td>4.1947</td>
<td>.0424880</td>
</tr>
<tr>
<td>134</td>
<td>1</td>
<td>3490.59</td>
<td>135</td>
<td>364.1313</td>
<td>9.5861</td>
<td>.0023847</td>
</tr>
<tr>
<td>234</td>
<td>1</td>
<td>41.92</td>
<td>135</td>
<td>364.1313</td>
<td>.1151</td>
<td>.7349270</td>
</tr>
<tr>
<td>125</td>
<td>2</td>
<td>329.04</td>
<td>270</td>
<td>121.7271</td>
<td>2.7031</td>
<td>.0688114</td>
</tr>
<tr>
<td>135</td>
<td>2</td>
<td>402.54</td>
<td>270</td>
<td>121.7271</td>
<td>3.3069</td>
<td>.0381177</td>
</tr>
<tr>
<td>235</td>
<td>2</td>
<td>66.09</td>
<td>270</td>
<td>121.7271</td>
<td>.5429</td>
<td>.5816947</td>
</tr>
<tr>
<td>145</td>
<td>2</td>
<td>12.14</td>
<td>270</td>
<td>92.5111</td>
<td>.1313</td>
<td>.8770323</td>
</tr>
<tr>
<td>245</td>
<td>2</td>
<td>104.77</td>
<td>270</td>
<td>92.5111</td>
<td>1.1325</td>
<td>.3237576</td>
</tr>
<tr>
<td>345</td>
<td>2</td>
<td>53.53</td>
<td>270</td>
<td>92.5111</td>
<td>.5787</td>
<td>.5613451</td>
</tr>
<tr>
<td>1234</td>
<td>1</td>
<td>104.10</td>
<td>135</td>
<td>364.1313</td>
<td>.2859</td>
<td>.5937542</td>
</tr>
<tr>
<td>1235</td>
<td>2</td>
<td>41.28</td>
<td>270</td>
<td>121.7271</td>
<td>.3391</td>
<td>.7127151</td>
</tr>
<tr>
<td>1245</td>
<td>2</td>
<td>96.57</td>
<td>270</td>
<td>92.5111</td>
<td>1.0439</td>
<td>.3535080</td>
</tr>
<tr>
<td>1345</td>
<td>2</td>
<td>143.86</td>
<td>270</td>
<td>92.5111</td>
<td>1.5551</td>
<td>.2130561</td>
</tr>
<tr>
<td>2345</td>
<td>2</td>
<td>98.48</td>
<td>270</td>
<td>92.5111</td>
<td>1.0646</td>
<td>.3463186</td>
</tr>
<tr>
<td>12345</td>
<td>2</td>
<td>70.27</td>
<td>270</td>
<td>92.5111</td>
<td>.7596</td>
<td>.4688407</td>
</tr>
</tbody>
</table>

(Order = 1; Rarity = 2; Likelihood = 3; Before/After = 4; Content = 5)
APPENDIX 4 - Stimulus materials, mean ratings and standard deviations from the pre-tests

Table of Contents

Appendix 4a - Materials, mean ratings and standard deviations from the pre-test
  establishing the perceived rarity of the evidential features .................................... 337

Appendix 4b - Materials, mean ratings and standard deviations from the pre-test
  establishing the perceived association between the evidential features............... 339
Appendix 4a - Materials, mean ratings and standard deviations from the pre-test establishing the perceived rarity of the evidential features

House Problem

Out of every 10,000 houses how many would you expect to

(a) have a swimming pool?
(b) have a garden?
(c) have a garage?
(d) be worth approximately £250,000?
(e) be worth approximately £40,000?
(f) have been built in 1994?
(g) have been built sometime between 1945 and 1985?
(h) be over 200 years old?

Engineer Problem

Out of every 10,000 civil engineers how many would you expect to

(a) drive a company car?
(b) earn approximately £25,000 a year?
(c) earn approximately £60,000 a year?
(d) work in the south-east of England?
(e) work in Plymouth?
(f) work in general construction?
(g) specialise in the construction of cantilever bridges?
(h) travel regularly to supervise projects abroad?
Car Problem

Out of every 10,000 cars bought new in 1988 how many would you expect to

(a) have a radio?  
(b) have a multi-disc CD player?  
(c) have leather upholstery?  
(d) have four doors?  
(e) have a current value of over £500?  
(f) have a current value of over £15,000?  
(g) never have had any mechanical problems?  
(h) have a top speed of over 90 mph?  
(i) have a top speed of over 165 mph?  
(j) have been recalled due to a design flaw?

<table>
<thead>
<tr>
<th>Mean Rating</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>92%</td>
<td>17%</td>
</tr>
<tr>
<td>11%</td>
<td>16%</td>
</tr>
<tr>
<td>13%</td>
<td>14%</td>
</tr>
<tr>
<td>58%</td>
<td>19%</td>
</tr>
<tr>
<td>76%</td>
<td>24%</td>
</tr>
<tr>
<td>12%</td>
<td>15%</td>
</tr>
<tr>
<td>12%</td>
<td>21%</td>
</tr>
<tr>
<td>78%</td>
<td>23%</td>
</tr>
<tr>
<td>6%</td>
<td>10%</td>
</tr>
<tr>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Spanish Villa Problem

Out of every 10,000 holiday villas in Spain, how many would you expect to

(a) have eight bedrooms?  
(b) cost £150 per week?  
(c) be within walking distance of a beach?  
(d) have been built in 1994?  
(e) cost £1,000 per week?  
(f) have two bedrooms?  
(g) be located over 100 miles from the sea?  
(h) have been built in the last twenty years?  
(i) to have a large balcony?  
(j) to have a sitting room?

<table>
<thead>
<tr>
<th>Mean Rating</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>14%</td>
<td>19%</td>
</tr>
<tr>
<td>51%</td>
<td>25%</td>
</tr>
<tr>
<td>67%</td>
<td>20%</td>
</tr>
<tr>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>8%</td>
<td>13%</td>
</tr>
<tr>
<td>40%</td>
<td>24%</td>
</tr>
<tr>
<td>9%</td>
<td>9%</td>
</tr>
<tr>
<td>84%</td>
<td>17%</td>
</tr>
<tr>
<td>52%</td>
<td>26%</td>
</tr>
<tr>
<td>83%</td>
<td>23%</td>
</tr>
</tbody>
</table>
Appendix 4b - Materials, mean ratings and standard deviations from the pre-test establishing the perceived association between the evidential features

House Problem

Out of every 1000 houses with a swimming pool, how many would you expect to

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean Rating</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) have a garden?</td>
<td>93%</td>
<td>12%</td>
</tr>
<tr>
<td>(b) have a garage?</td>
<td>94%</td>
<td>9%</td>
</tr>
<tr>
<td>(c) be worth approximately £250,000?</td>
<td>62%</td>
<td>21%</td>
</tr>
<tr>
<td>(d) be worth approximately £40,000?</td>
<td>14%</td>
<td>22%</td>
</tr>
<tr>
<td>(e) have been built in 1994?</td>
<td>13%</td>
<td>13%</td>
</tr>
<tr>
<td>(f) have been built sometime between 1945 and 1985?</td>
<td>58%</td>
<td>24%</td>
</tr>
<tr>
<td>(g) be over 200 years old?</td>
<td>7%</td>
<td>6%</td>
</tr>
<tr>
<td>(h) have a tennis court?</td>
<td>84%</td>
<td>16%</td>
</tr>
</tbody>
</table>

Engineer Problem

Out of every 1000 civil engineers who earn approximately £60,000 a year, how many would you expect to

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean Rating</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) drive a company car?</td>
<td>84%</td>
<td>16%</td>
</tr>
<tr>
<td>(b) work in the south-east of England?</td>
<td>47%</td>
<td>21%</td>
</tr>
<tr>
<td>(c) work in Plymouth?</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>(d) work in general construction?</td>
<td>36%</td>
<td>26%</td>
</tr>
<tr>
<td>(e) specialise in the construction of cantilever bridges?</td>
<td>8%</td>
<td>11%</td>
</tr>
<tr>
<td>(f) travel regularly to supervise projects abroad?</td>
<td>25%</td>
<td>27%</td>
</tr>
</tbody>
</table>
**Car Problem**

Out of every 1000 cars bought new in 1988 which have a top speed of over 165 mph, how many would you expect to

(a) have a radio?

(b) have a multi-disc CD player?

(c) have leather upholstery?

(d) have four doors?

(e) have a current value of over £500?

(f) have a current value of over £15,000?

(g) never have had any mechanical problems?

(h) have been recalled due to a design flaw?

<table>
<thead>
<tr>
<th>Mean Rating</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>18%</td>
</tr>
<tr>
<td>40%</td>
<td>33%</td>
</tr>
<tr>
<td>52%</td>
<td>31%</td>
</tr>
<tr>
<td>50%</td>
<td>26%</td>
</tr>
<tr>
<td>87%</td>
<td>23%</td>
</tr>
<tr>
<td>42%</td>
<td>28%</td>
</tr>
<tr>
<td>15%</td>
<td>20%</td>
</tr>
<tr>
<td>13%</td>
<td>21%</td>
</tr>
</tbody>
</table>

**Spanish Villa Problem**

Out of every 1,000 holiday villas in Spain which cost £1,000 per week, how many would you expect to

(a) have eight bedrooms?

(b) be within walking distance of a beach?

(c) have been built in 1994?

(d) have two bedrooms?

(e) be located over 100 miles from the sea?

(f) have been built in the last twenty years?

(g) to have a large balcony?

(h) to have a sitting room?

<table>
<thead>
<tr>
<th>Mean Rating</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>51%</td>
<td>28%</td>
</tr>
<tr>
<td>73%</td>
<td>15%</td>
</tr>
<tr>
<td>15%</td>
<td>9%</td>
</tr>
<tr>
<td>18%</td>
<td>25%</td>
</tr>
<tr>
<td>9%</td>
<td>7%</td>
</tr>
<tr>
<td>72%</td>
<td>27%</td>
</tr>
<tr>
<td>70%</td>
<td>25%</td>
</tr>
<tr>
<td>94%</td>
<td>8%</td>
</tr>
</tbody>
</table>


*Organizational Behaviour and Human Performance, 19, 43-55.*


