UNIVERSITY OF DERBY

RELEVANCE AND RATIONALISATION IN THE WASON SELECTION TASK

Erica Lucas

Doctor of Philosophy 2007

THE FOLLOWING HAVE NOT BEEN COPIED ON INSTRUCTION FROM THE UNIVERSITY

Figure 2.1 page 28

List of Contents

Chapter 3 – Experiments 4 to 7

Chapter 4 – General Discussion

Ball, L.J., **Lucas, E.J.,** Miles, J.N.V., & Gale. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology, 56A,* 1053-1077.

Lucas, E.J., & Ball, L.J. (2005). Think-aloud protocols and the selection task: Evidence for relevance effects and rationalisation processes. *Thinking and Reasoning, 11,* 35-66.

List of Tables

List of Figures

Abstract

Evans' (e.g., 2006) heuristic-analytic theory of the selection task proposes that card selections are triggered by relevance-determining heuristics, with analytic processing serving merely to rationalise heuristically-cued decisions. Evans (1996) provided evidence for the theory by setting up an inspection-time paradigm. He used computer-presented selection tasks and instructions for participants to indicate (with a mouse-pointer) cards under consideration. The theory predicts that longer inspection times should be associated with selected cards (which are subjected to rationalisation) than with rejected cards. Evans found support for this idea. Roberts (1998b) however, argued that mouse-pointing gives rise to *artefactual* support for Evans' predictions because of biases associated with the task format and the use of mouse pointing. In the present thesis all sources of artefact were eradicated by combining careful task constructions with eye-movement tracking to measure directly on-line attentional processing. Across a series of experiments good evidence was produced for the robustness of the inspection-time effect, supporting the predictions of the heuristic-analytic account. It was notable, however, that the magnitude of the inspection-time effect was always small. A further experiment separated the presentation of rules from associated cards to avoid possible dilution of the inspection-time effect arising from parallel rule and card presentation. However, the observed inspection time effect remained small. A series of experiments utilising think-aloud methods were then employed to test further the predictions concerning relevance effects and rationalisation processes in the selection task. Predictions in relation to these experiments were that selected cards should be associated with more references to both their facing and their hidden sides than rejected cards, which are not subjected to analytic rationalisation. Support was found for all heuristic-analytic predictions, even where 'select/don't select' decisions were enforced for all cards. These experiments also clarify the role played by secondary heuristics in cueing the consideration of hidden card values during rationalisation. It is suggested that whilst Oaksford and Chater's (e.g., 2003) information gain theory can provide a compelling account of our protocol findings, Evans' heuristic-analytic theory provides the best account of the full findings of the thesis. The mental models theory (e.g., Johnson-Laird & Byrne, 2002) fares less well as an explanation of the full dataset.

Acknowledgements

I am indebted to Dr. Linden Ball for his excellent supervision, his guidance and support during the completion of this thesis. He has been nothing but encouraging and patient, especially through the writing-up phase, and I am enormously grateful to him for it.

Special thanks must also go to the Applied Vision Research Unit at Derby (now at Loughborough) for the use of their eye-tracker and expertise. In particular, Alastair Gale for his useful comments and continued support, and Kevin Purdy and Mark Mugglestone for their technical expertise and time.

I would also like to thank all my friends and colleagues from the University of Derby, Nottingham Trent University and those elsewhere for all their support and encouraging words.

Lastly, I would like to thank my husband Spencer, my daughter Emelia and the rest of my family for their continued support throughout.

Chapter 1

The Psychology of Conditional Reasoning

1.1 Chapter Outline

This chapter provides an introduction to the psychological study of human reasoning, with a particular emphasis on issues relating to people's cognitive processes when reasoning about conditional statements of the form 'if...then'. Indeed, this thesis is primarily concerned with people's reasoning performance in relation to one specific conditional reasoning paradigm: the abstract version the famous four card *selection task* introduced by Peter Wason (1966). This task is described in detail below, but suffice it to say for now that it is based around an abstract indicative conditional rule and aims to assess people's hypothesis testing behaviour in relation to this rule. The task remains intriguing for two main reasons: first, most people get the problem wrong according to the logical standards of conditional inference; second, it has been devilishly hard to figure out why the logically incorrect responses that are made tend to be highly systematic. The present thesis will, therefore, continue a long line of important research in its attempt to further an understanding of these issues. Indeed, the whole thesis is predicated on the belief that we need to progress our theoretical accounts of the selection task before we can then determine what the selection task may be able to tell us about how people interpret and think about conditional rule-forms. Because the present thesis focuses on the selection task the present chapter will provide some particularly in-depth coverage of empirical findings and theoretical perspectives that surround the task.

This introductory chapter will be structured around the following key themes: (1) the importance of reasoning in everyday life; (2) the nature of deductive reasoning, conditional reasoning and propositional logic; (3) basic issues in the study of conditional reasoning, including the link between conditional reasoning and the rationality debate; and (4) the Wason selection task, its associated phenomena, and theories that have been proposed to explain these phenomena.

1.2 The Importance of Reasoning in Everyday Cognition

The ability to make inferences that go beyond given information is commonly viewed as being fundamental to human intelligence (e.g., Johnson-Laird & Byrne, 1991; Oakhill & Garnham, 1994; Evans, Newstead, & Byrne, 1993). Deriving inferences is, itself, a core element of the reasoning and decision-making that makes humans so different to other animals (cf. Stanovich, 1999; 2004). We use reasoning to learn concepts and principles and to generalise what we have learnt from one situation to another (e.g., Holland, Holyoak, Nisbett & Thagard, 1986). Without reasoning we would not be able to understand utterances in language (e.g., Sperber & Wilson, 1995) or engage in everyday problem solving (e.g., Holyoak & Thagrad, 1995). Intriguingly, much reasoning seems to take place automatically and without conscious awareness. For example, without realising the complex processing that is taking place, we are able to talk to people and understand what they are saying, we are able to attend to problems and draw on relevant information from our memory to try to solve them, and we can engage in sophisticated planning so as to imagine the shape and form of future events – both in the short-term and the long-term.

However, our reasoning abilities, whilst often effective in many practical contexts, are not infallible, and we do make mistakes. We might solve a problem, generate a forecast of the future and plan a course of action all based on a faulty judgement in the first place, and mistakes like this can be costly and may be also be dangerous. Some examples of this in recent history include: (1) the case of the scientists who judged that the Chernobyl nuclear reactor was safe during a routine test, when, in fact, a danger state was imminent (26th April 1986, Medvedev, 1991) and (2) the pilot who turned off the wrong engine to isolate an engine fire that led to a passenger aircraft crashing near the village of Kegworth in Derbyshire (8th January 1989; Trimble, 1990). In business situations, too, a wrong decision may mean putting a company out of business, whilst in medical research it might mean inaccurate claims about a drug.

It is important to note, though, that a bad outcome does not necessarily mean poor thinking, as there may be a lack of information available to the reasoner at the time. In the same way, a good outcome may not necessarily reflect sound reasoning (as we will see later, it might be that a decision produces a response that coincides with the logically correct choice). The significance of all these points, then, is that they demonstrate how vitally important it is to study human reasoning in order to understand the nature of systematic errors (or biases) that influence the inferences that people draw. It is only by deriving a clear understanding of the nature and causes of such biases that an attempt can be made to reduce their impact through education, training or other remedial approaches (cf. Baron, 2000; Evans, 1989; Stanovich, 2004).

1.3 Deductive Inference, Conditional Reasoning and Propositional Logic

There are a number of different *types* of inference that are commonly discussed in the reasoning literature, and a particularly common distinction is between an *inductive* inference and a *deductive* inference (though see C.S. Pierce [Weiner, 1958] for an intriguing discussion of so-called *abductive* inference, which is aimed at finding the best explanation for surprising or anomalous observations). An inductive inference is one that generalises from specific everyday experiences toward encompassing rules or laws that capture observed regularities (e.g., that all metals expand when they are heated). An inductive inference adds information and involves reasoning that progresses from the specific to the general. An inductive inference, therefore is not necessarily true and may well be falsified by a future observation (e.g., the identification of a metal that fails to expand when it is heated).

A deductive inference, on the other hand, is one where the conclusion that is drawn *necessarily* follows from the information given (i.e., it is not possible for the given information to be represented in a way that would render the conclusion false). No new information is added in a deductive inference, that is, reasoning occurs from the general to the specific. Although an inductive inference is not logically sound, both types of inference are important in our everyday thinking about the world. However, it is deductive inference that is the focus of interest in the present thesis. In the past, research into deductive reasoning has been dominated by so-called *normative* theories that determine whether inferences are correct or incorrect according to formal standards of effective thinking. In the case of deductive reasoning, the normative theory that embodies a priori criteria for determining the validity of inferences is that of *logic*. Most of the work on deductive reasoning involves the normative framework afforded by *propositional logic* – especially in the case of *conditional reasoning*, which forms the focus of this thesis.

When conditional reasoning tasks are used in experiments, correct or incorrect answers are categorised by the logical axioms of the propositional calculus (e.g., see Hodges, 1993). Propositional calculus uses the terms *p* and *q* to express the atomic propositions that make up a statement, and the logical operator '*if.....then*' to relate the atomic propositions together (see Evans, 1982). As such, statements taking the basic form '*if p then q*' are traditionally how conditionals have been studied in the psychological laboratory. Propositional calculus also involves a number of operators in addition to *if....then*, such as *not*, *and*, *or* and, less commonly, *if and only if*. According to propositional calculus, a proposition can only be assigned one of two logical values: 'true' or 'false'. There can be no in-between values – and therefore no uncertainty about the truth status of a propositional term.

In order to work out the relation between a logical operator and associated propositions a syntactic formalism referred to as 'truth table analysis' can be used (see Johnson-Laird & Byrne, 1991). Truth table analysis enables a decision to be made as to whether a statement involving propositional terms and one or more logical operators is valid or invalid. When we have two propositions, *p* and *q*, these can each be either true or false, and when combined together there are four possible states of affairs, as depicted in the following truth table:

p	***q***
T	T
T	F
F	T
F	F

From this truth table it can be seen that both p and q can be true, p can be true when q is false, p can be false when q is true, and, finally, both p and q can be false. If we now consider a real-world conditional of the form *if p then q*, such as 'if the switch is pressed the kettle boils' (where p = the switch is pressed and q = the kettle boils) we can establish the following truth table for the conditional:

		conditional
p	***q***	***if p then q***
T	T	T
T	F	F
F	T	T
F	F	T

First, if both p and q are true then *if p then q* is true. If we know that the switch is pressed and that the kettle boils, then we can be sure that 'If the switch is pressed then the kettle boils'. However, if p is true and q is false (the switch is pressed and the kettle does not boil) then the assertion *if p then q*, 'If the switch is pressed then the kettle boils', is false. Moving on to the next two instances, if p is false and q is true (although the switch is not pressed the kettle still boils) then the statement 'If the switch is pressed then the kettle boils' is true. In this case we do not actually know whether the statement is true or not, or we are not happy saying so, but according to logic it has to be either true or false. The assertion has to remain true because there are no grounds for it to be false – something else could have caused the kettle to boil such as putting it on the hob. Therefore the assertion has to be true. Also, when both p and q are false (the switch is not pressed and the kettle does not boil) the assertion is also true according to propositional logic.

There are four inferences associated with conditionals: Modus Ponens (MP), Modus Tollens (MT), Denial of the Antecedent (DA) and Affirmation of the Consequent (AC). When we consider these inferences in terms of a truth table for a conditional inference we can see that only MT and MP are valid inferences, whereas DA and AC are invalid inferences, otherwise known as *fallacious inferences*. For example with the conditional 'if the switch is pressed then the kettle boils' (*if p then q*) and we are told that the switch is pressed (*p* is true), then the correct conclusion is 'the kettle boils' as there is only one line in the table where *p* is true and where *if p then q* is true (*q* is also true). The inference is therefore valid. The same can be said for the MT inference. If we are given the conditional 'if the switch is pressed then the kettle boils' and then told that 'the kettle does not boil' then we can only infer from the truth table that *p* is also false, i.e. 'the switch is not pressed'. Therefore the inference is also valid.

The other two inferences are more difficult. In terms of the AC inference if we are given the statement 'if the switch is pressed then the kettle boils' (*if p then q*) and are told that 'the kettle boils' (*q* is true), then when we consider this in terms of the truth table we notice that there are two lines that fit this situation and in one case *p* is true and in the other *p* is false. Because we have no clear answer in terms of logic it means that we cannot conclude anything from this statement, therefore the inference is invalid. The same can be said for the fourth inference, DA. Given the statement 'if the switch is pressed then the kettle boils' (*if p then q*) and 'the switch is not pressed' (*p* is false), we note again that there are two lines in the truth table where *if p then q* is true and *p* is false. We now have the same situation as we had with *q* in the AC inference where in one line *q* is false and one line *q* is true. No conclusion can be drawn and so the DA inference is invalid.

Experiments on conditional reasoning may involve the presentation of truth table cases and the requirement for participants to judge whether the presented case corroborates the given rule or contradicts it (in some experiments an 'irrelevant' response is also permitted so that people can register their belief that the case has no bearing on the given conditional). An alternative to this 'truth-table evaluation task' is the 'truth-table construction task', where participants are requested to generate confirming or falsifying instances for the given conditional (see Evans, 1998b, for a review of key phenomena associated with truth-table construction and evaluation tasks). An alternative approach to examining conditional reasoning is for the MP, MT, AC and DA arguments outlined above to be presented to participants with a requirement for them to make a validity judgement (see Evans et al., 1993, and Evans, 1998b, for reviews). This 'conditional-inference paradigm' may either involve presenting the participant with a choice of conclusion and asking them to choose the valid statement from the list of possible conclusions (an evaluation task), or it may be done by asking them to produce their own conclusion (a generation task).

Yet another approach to examining conditional reasoning involves presenting participants with a reasoning task that requires them to understand logical relationships by asking them to consider and evaluate alternative hypotheses. One famous example of this is the four-card selection task of Peter Wason (1966), which is the task that is the focus of the research presented in this thesis. We turn to a detailed overview of the selection task and its many variants in Section 1.5 below, after first considering some of the basic issues to have emerged in the study of conditional reasoning.

1.4 Key Issues in the Study of Conditional Reasoning

Traditionally, reasoning researchers have often been preoccupied with how *accurate* participants' inferences are according to the normative standard of formal logic. Studies that have examined this issue have tried to determine when errors are made, and whether errors are random or systematically biased by features of the task. Errors have been interesting to researchers because it is thought that if we can discover the conditions under which errors are produced then we can learn something about underlying cognitive processes. Debate in the reasoning literature has tended to focus on how biases come about, and whether they derive from a rational attempt to reason or from some other source. These alternative views of biased reasoning can be illustrated if we look at the case of Modus Tollens conditional inference introduced earlier. If we consider the statements:

If it is a square then it is orange
It is green

It logically follows that it is *not* a square. If it were a square then it would have to be orange. Although a majority of people appear to make this valid Modus Tollens inference it is often reported that a substantial minority of participants say that no conclusion can be drawn (see for example Evans, 1977, where 75% of participants drew the inference). The interest then is why some people readily draw this valid inference whilst others do not. The situation can be complicated further by adding a negative in the first (major) premise, for example if we say:

If it is not a square then it is orange
It is green

In this case we can still make the Modus Tollens inference but it leads to a different conclusion, that it *is* a square because if it were not it would have to be orange. Evans (1977) showed, however, that this change in the major premise leads to a huge change in participants' responding so that now only 12% of participants draw the Modus Tollens inference. The presence of the negative can be described as a biasing factor, and it is therefore clear that to understand inferences in reasoning it is not just a case of understanding whether people can draw an inference, it is also critical to determine when they fail to do so and under what circumstances.

The above statements are examples of an *abstract* reasoning problem, that do not have any connection with prior knowledge. If we start to add real world content (or 'thematic' content) to a reasoning task then things start to get even more complicated. Although many experiments have asked participants to ignore their prior knowledge or beliefs in reasoning, these experiments have shown that people find this incredibly hard to do (Evans et al., 1993). If we look at the statement:

If it is a square then it is orange
It is a triangle

according to logic we can not infer anything about the colour of this shape. Drawing the conclusion that it is not orange is committing the denial of the antecedent fallacy. As Evans (1977) has shown, with the basic kind of conditional statement depicted above (i.e., affirmative and abstract), simply adding a negation to a term can strongly affect people's responding. Interestingly, however, when we add thematic content to a statement like this, the effect of the negative disappears. It is now context that plays a critical role. Ellis (1991, Experiment 5) for example, embedded problems in short scenarios such as:

If the truck is heavier than the legal limit, then the alarm bell will ring.
The truck is under the legal limit

and found that most of the participants drew the denial of the antecedent fallacy by inferring that the alarm bell would not ring. When the materials used arbitrary universal claims however, such as:

If the student is doing Economics, then he is a socialist

they only drew the inference around 50% of the time. So it seems that pragmatic factors that involve prior knowledge have a big effect on the way we reason.

Evans (1991) proposes that research into reasoning has been motivated by three separate but interconnected issues. He calls these the competence question (by what mechanism can participants reason out the solution to logical problems?), the bias question (what factors cause systematic errors and biases in reasoning and what do such biases tell us about the nature of reasoning processes?), and the content question (what features of the task content and context affect the ability of participants to reason the solution to logical problems and what does this tell us about the nature of the reasoning process?). We will come back to these issues later when we discuss phenomena that occur on the selection task, but we pick up on how these issues play out in relation to the so-called 'rationality debate'.

Experiments reported in the literature on human deductive reasoning with statements such as conditionals have led to two main conclusions: (1) that observed *competence* on reasoning tasks is often low as measured according to the 'gold standard' of normative logic; and (2) that *performance* is often inconsistent across different reasoning tasks and is susceptible to a range of experimental variables, such as those associated with manipulations of problem content or associated context (Evans, 1989). These conclusions have been fundamental to the emergence of the 'rationality debate', in that performance, according to logicality, has often been seen as an indicator of rationality in reasoning. This inconsistent pattern of performance on reasoning tasks contradicts what we know about human intelligence, that is, that humans are generally seen to be rational and intelligent entities (e.g., Cohen, 1981).

Cohen (1981) in fact claims that experiments on deductive reasoning will never be able to demonstrate human irrationality. He argues that how people perform on laboratory-based reasoning tasks should not be confused with underlying competence because these tasks are, by necessity, artificial and unrepresentative of real-world problems. Indeed, it might even be the case that participants are reasoning according to logics other than the propositional calculus. Since Cohen's (1981) influential paper there have been a variety of criticisms directed at the reasoning and decision-making literature about the external validity of the reported research. It has been suggested that researchers have been too quick to generalise from laboratory studies to the real-world. However, Evans and Over (1996) suggest a different way of conceptualising rationality. They distinguish between $rationality_1$, where people act in a reliable way in order to achieve their real-world goals, and $rationality_2$, where people act for reasons sanctioned by a normative theory. Within the reasoning literature, then, performance on reasoning tasks can be seen to be irrational according to normative theory (failure to be $rational_2$), but still be rational with regards to achieving real-world goals ($rational_1$).

The distinction between $rationality_1$ and $rationality_2$ is not just a way of defining and thinking about rationality. Evans and Over (1996) go further and make an actual *psychological* distinction between implicit and explicit *systems* of cognition that they argue map onto these two types of rationality (for related dual-system views of cognition see Reber, 1993; Stanovich, 1999, 2004; Sloman, 1996). For Evans and Over (1996), the 'implicit' system involves the application of tacit knowledge and procedures obtained through prior learning and experience. The 'explicit' system, in contrast, is characterised as involving conscious thinking and embodies mechanisms whereby we model hypothetical states and consequences. Both the implicit and the explicit system are constrained: the implicit system because of its reliance on learning from past experiences; the explicit system because of its limited capacity that is bounded by working memory constraints. Rationality, then, is restricted generally – and rationality$_2$ more so than rationality$_1$, because the explicit system relies on the implicit system to pass relevant information on to it. Evans and Over (1996) therefore propose that we do have the ability to be both rational$_1$ and rational$_2$. Although we sometimes use rationality$_2$ in everyday life, most of our routine decisions involve the application of rationality$_1$.

The rationality$_1$ versus rationality$_2$ distinction affords a neat resolution to the rationality debate, although it is not without its critics (see the accompanying replies to Evans & Over's, 1996, proposals in the special issue of the journal *'Current Psychology of Cognition'* in 1997). Indeed, in his more recent discussions of reasoning processes, Evans (e.g., 2006) has essentially dropped any mention of the rationality$_1$/rationality$_2$ distinction, instead simply invoking a 'dual processing' view of cognition, which sits at a lower level of theoretical abstraction than an all-encompassing dual-system framework. This affords the advantage of reducing the amount of 'theoretical baggage' that comes along with the dual-system perspective, thereby avoiding many of the challenges that can be lodged against the broader scheme (see Evans, 2006). Evans' dual-process views of reasoning will be discussed in more detail below in the context of his current heuristic-analytic dual-process theory of reasoning (Evans, 2006) that incorporates 'three principles of hypothetical thinking'.

Unlike Evans' (2006) move to a lower-level dual-process characterisation of reasoning and issues relating to rationality, however, we note that other authors, most notably Stanovich (2004), remain deeply wedded to a generalised dual-system perspective as a way to account for conflicts between our evolutionary-older reptilian minds whose primary goal is that of the replication of the species (cf. Dawkins, 1976; 1996) and our more modern capacities for self-reflective thought and autonomous control over our biological programming (i.e., analytic rationality). According to

Stanovich (1999, 2004), System 1 developed early in evolutionary prehistory and forms the universal basis for all animal cognition. In this system, knowledge is retrieved and applied rapidly and automatically and it is therefore described as implicit and unconscious. It does not correlate with individual differences in working memory capabilities or measured general intelligence (see also Stanovich & West, 2000) and appears to operate through associative neural networks. The processes carried out in System 1 reflect high-level functioning of universal abilities such as language comprehension. In contrast, System 2 is said to have evolved late and is unique to human beings. Knowledge is retrieved much slower and sequentially and so is described as explicit, conscious and analytic. It operates through working memory and does correlate in its efficiency with measured general intelligence (Stanovich & West, 2000). Processes carried out in this system reflect heritability of cognitive capacity measured by IQ, which vary according to inherited genes.

1.5 The Wason Selection Task Phenomena

Having outlined some of the fundamental issues surrounding the study of conditional reasoning—including those of deductive competence, bias and rationality—we now turn our attention to the conditional reasoning paradigm, the Wason selection task, that formed the focus of the empirical research that will be presented in this thesis. We start off by overviewing the nature of the task and the way that it links with concerns about people's understanding of conditional statements. We next examine key phenomena that have been uncovered over the past 40 years of research that has made use of the basic selection task as well as variants of it. Finally this section will provide an in-depth critical discussion of the main theories that have been proposed to account for people's performance on the task.

The selection task that was originally developed by Wason (1966) involved presenting participants with four cards, derived from a larger pack of similar cards that each have a letter on one side and a single-digit number on the other side. The four cards that are presented only show their facing value (see Figure 1), which in the present case are 'A', 'D', '3' and '7'. These values are often referred to as the *p*, *not-p*, *q* and *not-q* cases for a rule of the form '*if p then q*'. Participants are then given a rule that they are told applies to the four cards, and their task is to decide which card(s) need to be turned over to determine whether the rule is true or false. The rule is as follows: 'If there is an A on one side of the card then there is a 3 on the other side of the card'.

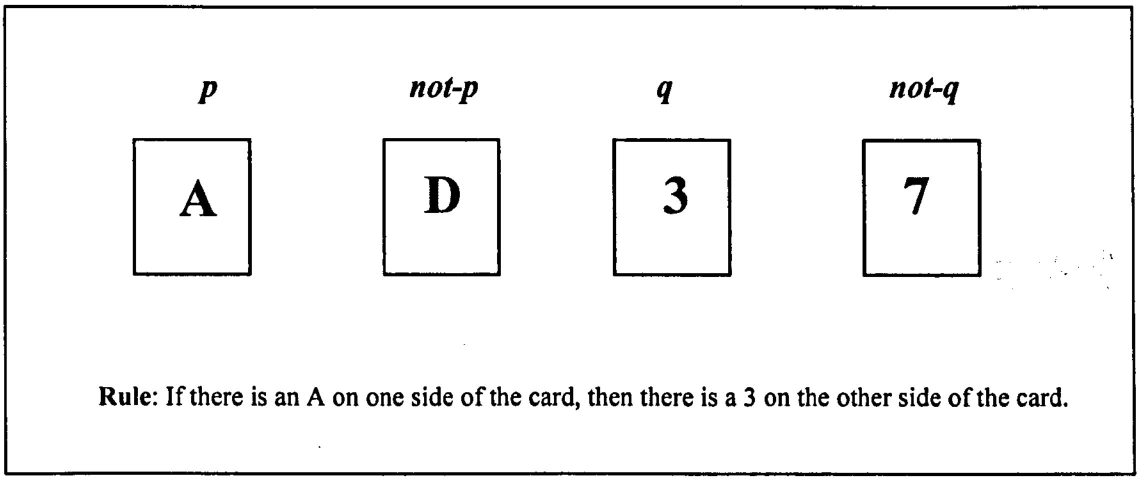

FIGURE 1.1

The standard abstract version of the Wason selection task

The logically correct response to this problem is to turn the A (*p*) and the 7 (*not-q*) cards as these are the two cards that could potentially provide a letter-number combination that could show the rule to be false. This can be seen more clearly if we project the logical consequences of turning each card. The A card may have either a 3 on the back, which conforms to the rule, or another number which contradicts the rule, therefore it has to be turned over. The D card does not have to be turned over as the rule has no implication for a card with a D on it. The 3 card may have an A on the back but if it does not the rule is not contradicted as there is no claim that there is an A on the back of a 3. There is then, no need to turn this card. Lastly, the 7 card might have an A on the back, and this would then be a case of an A with a number that is not a 3. As this clearly contradicts the rule, the 7 card needs to be turned over.

Despite its apparent simplicity, studies have shown that fewer than 10% of people make the logically correct choices A (*p*) and 7 (*not-q*), with the majority of people choosing either the A (*p*) and the 3 (*q*), or just the A card (see for example Wason, 1968, 1969; Wason & Johnson-Laird, 1970). Indeed, two logical errors that persist on the task are: (1) the selection of the *q* card; and (2) the failure to select the *not-q* card.

The wealth of literature on the selection task over the last four decades has, in the main, come from researchers' attempts to explain these logical mistakes. This effort has, in fact, led researchers in two quite different directions. One direction has involved attempts to define what features of the task make it so hard, and, leading on from that, what task manipulations will make it easier. As a result, there is a body of empirical research on a number of different task phenomena. These phenomena arise from: (1) changing the form of the rule; (2) changing the content of the task; (3) changing the context of the task, and (4) changing the task so that participants can envisage alternatives explicitly. Key aspects of the empirical research addressing these phenomena will be considered next. The other direction that researchers have taken has been to prioritise the formulation of theoretical explanations of the behaviour observed on the task. This theoretical body of research will be dealt with later on in this section after reviewing empirical findings concerning key selection-task phenomena. It is of course, important to note here that the empirical phenomena have informed theory development, and theories, likewise, have informed the range of empirical phenomena that have been examined. This separation of empirical findings from the theoretical issues is, therefore, at least in part aimed at managing the complexity of the vast literature surrounding the selection task, though it remains that some researchers have taken a more empirical stance on the task whilst others have been more concerned with theory development.

1.5.1 Changing the form of the rule

Wason (1966) originally suggested that people were showing some sort of confirmation or verification bias on the abstract selection task, that is, they were trying to prove the conditional rule true by choosing the *p* and *q* cards instead of trying to demonstrate the potential falsity of the rule according to normative logic. However, subsequent research by Evans and Lynch (1973), which introduced negations into the conditional rule, showed that participants exhibit a systematic error known as *matching bias* whereby they simply choose the cards that are named in the given rule. In order to understand matching bias, it is necessary to define the logical status of each of the four cards because reference to card choices as *p, not-p, q* and *not-q* becomes ambiguous once negations are introduced. The cards are therefore usually referred to as True Antecedent (TA), False Antecedent (FA), True Consequent (TC) and the False Consequent (FC). On the affirmative rule '*If p then q*', these definitions relate to the *p, not-p, q* and *not-q* card respectively. When using a negated conditional such as, '*If p then not q*', however, the negative consequent reverses the correspondence between the negation in the instance and its logical status so that TC is now represented by *not-q* and FC by *q* (refer to Table 1.1 below to see this more clearly for the logical cases across all rules in the negations paradigm).

Permuting negatives through the conditional rule allows the separation of confirmation bias and matching bias. If confirmation bias was being exhibited then participants should choose the TA card and the TC card every time regardless of the presence of negatives in the rule. However, if they are matching they should make the *p* and *q* choices over the *not-p* and *not-q* choices on each logical case, i.e. the cards mentioned in the rule. Evans and Lynch indeed found that people matched consistently those cards that were mentioned in the rule, despite the existence of negatives.

TABLE 1.1

Negations paradigm applied to the selection task showing matching antecedent cases (*p*) and matching consequent cases (*q*) across all four rules

	TA	FA	TC	FC
If p then q If there is an A, then there is a 3	A (*p*)	D (*not-p*)	3 (*q*)	7 (*not-q*)
If p then not q If there is an A, then there is not a 3	A (*p*)	D (*not-p*)	7 (*not-q*)	3 (*q*)
If not p then q If there is not an A, then there is a 3	D (*not-p*)	A (*p*)	3 (*q*)	7 (*not-q*)
If not p then not q If there is not an A, then there is not a 3	D (*not-p*)	A (*p*)	7 (*not-q*)	3 (*q*)

Matching bias appears to be a robust phenomenon that arises across a variety of different tasks. It has been demonstrated in selection tasks that use abstract rules with various connectives, including: *if p then q*, *q if p*, *p only if q* and *there is not both p and q* (e.g., Evans, Clibbens, & Rood, 1996; Evans, Legrenzi, & Girotto, 1999). Furthermore, Roberts (2002) has shown that matching occurs in conditional selection tasks using categorical rules of the form *all p have q*. In addition to this, matching bias has been observed on other types of task such as the truth table construction task, where people are asked to *construct* instances that verify, falsify or are irrelevant to a given conditional or categorical rule, and on the truth table evaluation task where they are asked to *identify* instances that verify, falsify or are irrelevant to a given conditional or categorical rule (e.g. Evans, 1998b; Evans et al., 1999).

However, it seems somewhat unclear as to whether matching bias can be extended to rules such as disjunctions. Whilst some evidence from truth table tasks suggests such generality (Evans et al., 1999; Evans & Newstead, 1980), evidence from disjunctive selection tasks is extremely inconsistent. Van Duyne (1973, 1974) observed no effect of matching bias on disjunctive rules. Evans et al., (1999) observed a weaker matching bias effect compared with other connectives and Krauth and Berchtold-Neumann (1988) found no matching bias effect with inclusive disjunctives but an effect with exclusive disjunctives. A more recent study by Roberts (2002) that set out to investigate this disparity revealed an inverted effect, so that there were fewer matching than mismatching selections. The concept of matching bias and its underlying determinants will be a constant theme within this chapter.

1.5.2 Changing the content of the task

Many reasoning tasks have been presented with abstract content with the idea that reasoning with abstract content allows us to tap into some kind of 'pure' reasoning ability. However, as Manktelow (1999) suggests, this assumption is problematic for two reasons. First, reasoning with abstract materials has been shown to be subject to a number of biases, for example, the matching bias that is observed on the selection task as described above. Second, it may be that the form and content of the task cannot be separated, so that *how* you think cannot be separated from *what* you think. This second point contributed to the shift in perspective to using the task content to explain the task more fully. Indeed, research with the Wason selection task accounts for the majority of empirical research that has examined content effects in reasoning tasks, outside of the 'belief bias' effect with categorical syllogisms (e.g., Ball, Phillips, Wade, & Quayle, 2006; Evans, Handley, & Harper, 2001; Morley, Evans, & Handley, 2004).

Within early selection-task research it seemed that altering the content of the task by using realistic or thematic materials significantly improved people's performance. Wason and Shapiro (1971) reported an early study exploring this 'thematic facilitation effect'. The materials they used are now well known as the 'Towns and Transport' problem. Participants were given the rule *'Every time I go to Manchester I travel by car'*. The cards that were shown each represented a journey made by the experimenter with the destination on one side and the type of transport used to get there on the other. The four cards were 'Manchester' (*p*), 'Leeds' (*not-p*), 'Car' (*q*) and 'Train' (*not-q*). As in the abstract version of the task, participants were required to turn over the card(s) in order to decide whether the experimenter's claim was true or false. Wason and Shapiro found that significantly more people turned over the logically correct answers Manchester (*p*) and train (*not-q*), compared with people who were given the abstract version of the task.

Facilitation effects were also found on a task presented by Johnson-Laird, Legrenzi and Legrenzi (1972) that came to be known as the 'Postal Rule' problem. In this thematic version of the selection task, participants were asked to imagine that they were a postal worker testing the rule: *'If a letter is sealed then it has a 50 lire stamp on it'*. Participants were then shown a series of envelopes (rather than cards). The envelopes were either lying face down so participants could see whether or not the letter was sealed but could not see the value of the stamp on the front, or they were lying face up so it was possible to tell what stamp was on the letter but not whether it was sealed (see Figure 1.2). Most of the participants correctly selected the sealed envelope (*p*) and the letter stamped less than 50 lire (*not-q*).

FIGURE 1.2

Materials from the postal rule problem (Johnson-Laird, Legrenzi, & Legrenzi, 1972)

Despite the apparent 'thematic facilitation effects' demonstrated in these pioneering studies - as well as a few early replications of such facilitated conditional reasoning with very similar tasks - the view that these experiments were revealing and enhanced logicality purely as a function of realistic rule content, did not go unchallenged. For example, Manktelow and Evans (1979) set out to determine whether or not the matching bias established with the abstract task stood up to thematic content. They used arbitrarily realistic rule content (e.g., '*If I eat haddock then I drink gin*') and the full negations paradigm (one of the few studies with thematic material to do so). A control group was given the standard abstract task. The results deriving from their 'Food and Drinks' rules were surprising in that they failed to find the facilitation effect found in previous studies. The matching bias effect, however, was present in both the experimental and the control groups. Manktelow and Evans then directly replicated the Towns and Transport rule and again found no facilitation, instead revealing a matching effect. They suggested that the facilitation found on the Postal Rule by Johnson-Laird et al. (1972) was actually a memory effect because this rule related directly to a real rule experienced by the participants that had been in force in England prior to the 1972 study.

Griggs and Cox (1982) attempted to clarify the issues surrounding thematic contents in the selection task and named Manktelow and Evans' explanation as the 'memory-cueing' hypothesis. In a series of experiments, they were eager to show that rules using thematic content only facilitated correct selections when there was an opportunity for participants to use 'prior experience' to help solve the problem. They firstly replicated the Towns and Transport problem and, as predicted, found no facilitation due to the arbitrary nature of the materials and the lack of any relation to a 'real' rule. Next, they repeated the Postal Rule problem. Griggs and Cox predicted that this problem would produce no effects as their participants were from Florida and would have no experience of the rule. Again, as predicted, they found no facilitation. They then introduced a new problem known as the 'Drinking Age' rule problem. The rule within this problem was based on an actual rule used in Florida at that time. The four cards represented a person drinking in a bar with a type of drink on the other side of the card. The cards used were: drinking beer (*p*), drinking coke (*not-p*), 22 years of age (*q*) and 16 years of age (*not-q*). The rule used was: *'If a person is drinking beer then that person must be over 19 years of age'*. The whole problem was set in context with use of a short scenario that asked participants to imagine they were police officers checking a bar to ensure that under-age people were not drinking alcohol. As expected, very high facilitation effects were produced with most participants choosing the beer drinker (*p*) and the person aged 16 years old (*not-q*).

Later studies attempted to find out exactly what caused the facilitation on this problem and this resulted in variations in the context surrounding the violation-checking scenario as well as the instructions accompanying the task. For example, Pollard and Evans (1987) showed that the removal of the police officer scenario had detrimental effects on performance, with participants performing at abstract task level. Evans, Newstead, and Byrne (1993) suggested that memory-cueing can be used as the basis of a general 'availability' theory of content effects. The suggestion is that specific information and associations from memory influence the selections made in thematic versions of the task. If facilitation is achieved it is due to correct responses being made 'available' from prior experience and not necessarily down to a process of logical reasoning. Some of these ideas (i.e., that certain information is 'relevant' in the task content) assisted the development of the heuristic-analytic theory by Evans (e.g., 1984, 1989, 1998b; Evans & Over, 1996; 1997) which will be considered in a later section below.

It soon became clear, however that the specific process of memory-cueing was not able to explain all facilitation effects with thematic materials in the selection task. Studies have shown that you do not have to be able to retrieve information from your own experience in order to solve the task. For example, an experiment reported by D'Andrade (presented by Rumelhart, 1980) gave a scenario where participants had to imagine they were a manager in a Sears store where they had to check to see if sales receipts complied with rules. The rule was '*If a purchase exceeds $30 then the receipt must be approved by the departmental manager's signature*'. The cards were constructed so that they were receipts that showed the amount spent on one side and the presence or absence of the manager's signature on the other side. Cards were $50 (*p*), $30 (*not-p*), a signed receipt (*q*) and an unsigned receipt (*not-q*). Most people were able to solve this task correctly, selecting the receipt for over $30 (*p*) and the unsigned receipt (*not-q*). However, very few people would have had direct experience of the particular content of the task. This can also be seen in Johnson-Laird et al.'s (1972) original study in that some of the British participants were given the postal rule task that included Italian stamps (i.e. they had not necessarily had experience of posting letters in Italy).

At the end of the 1980s, then, it seemed far from clear what it was about certain tasks that facilitated or did not facilitate performance on the selection task. However, as the memory-cueing hypothesis lost favour, the enduring point was that facilitation might be due to a more basic feature of the task, such as the understanding of *social regulations*. The proposal, then, was that it may not matter so much whether reasoners have had direct experience of a particular scenario for selection-task facilitation; instead what is critically important is that reasoners have had experience of dealing with rules and regulations *in general*, and therefore know what should happen when a rule has been violated. It is this recognition by researchers that 'regulations' may be a key element of facilitated responding on thematic selection tasks that led to subsequent work on 'deontic reasoning' that is described in the next section.

1.5.3 Changing the context of the task to a deontic framework

Many of the thematic versions of the selection task discussed in the previous section involved the participant having to look for potential violations of the rule and not having to determine the truth or falsity of a rule statement as in the abstract version of the task. Since the late 1980s it has become increasingly apparent that there are, in fact, two distinct *types* of selection task that have emerged over the course of its history: the 'indicative' task and the 'deontic' task. The type is determined by the kind of conditional that is used in the task as well as the surrounding instructional context and scenario. Indicative tasks concern fact-based rules that may be true or false such as '*If I go to work, then I travel by train*'. Of course, the main example of the indicative form of the task is the original abstract version; however, we can see from the review of the studies in the previous section that there are some thematic tasks in the literature that also take the indicative form (e.g., the Towns and Transport problem and the Food and Drink problem). In contrast to indicative selection tasks, deontic versions of the task involve rules that are stated in order to direct people's behaviour such as '*If you travel by train then you must buy a ticket*'. Because deontic rules have a directive function (e.g., serving to encourage legal, moral, social, organisational or prudential behaviours) they may be obeyed or disobeyed. Indeed, the violation-checking scenario that typically contextualises such rules in selection task studies is specifically geared towards tapping the occurrence of people disobeying the rule. Such deontic rules along with violation-checking scenarios appear in both the Postal Rule problem and the Drinking Age problem.

Deontic rules were not *explicitly* tested in selection tasks until Cheng and Holyoak's pioneering study in (1985) that was conducted to test their 'pragmatic reasoning schemas' theory of facilitated selection-task performance. Instead of suggesting that we have knowledge of particular rules based on prior experience, as the memory-cueing hypothesis suggested, Cheng and Holyoak (1985) proposed that we have knowledge about rules and regulations generally. Their theory advanced the idea that we all possess *knowledge schemas* (i.e., packages of knowledge regarding specific domains like social regulations) and these schemas provide rules for thought and action. Importantly, Cheng and Holyoak argued that even without knowledge of a particular deontic rule, it should still be possible to use scenario-based cues to evoke a permission schema for an unfamiliar deontic rule, and thereby promote facilitated selection-task performance.

Cheng and Holyoak gave participants the postal rule task (described earlier), involving the rule '*If an envelope is sealed then it must have a 20 cent stamp on it*', and what they called the 'cholera task' involving the rule '*If a passenger's form says ENTERING on one side then the other side must include cholera amongst the list of diseases*'. This problem contained the following scenario (Cheng & Holyoak, 1985; pg 400):

> "You are an immigration officer at the Intentional Airport in Manila, capital of the Philippines. Among the documents you have to check is a sheet called Form H. One side of this form indicates whether the passenger is entering the country or in transit, while the other side of the form lists the names of tropical diseases. You have to make sure that if the form says 'ENTERING' on one side, then the other side includes cholera among the list of diseases. Which of the following forms would you have to turn over to check? Indicate only those that you need to check to be sure".

Participants then had four cards to choose from: one with '*Transit*', one with '*Entering*', one with '*Cholera, typhoid, hepatitis*' and one with '*Typhoid, hepatitis*'. Participants received two versions of each task, one with no rationale (the standard task) and one with a rationale. For the postal rule version the rationale was as follows (Cheng & Holyoak, 1985; pg 400):

> "The rationale for this regulation is to increase profit from personal mail, which is nearly always sealed. Sealed letters are defined as personal and must therefore carry more postage than unsealed letters".

In the cholera version instead of the form listing the tropical diseases, it listed inoculations the passenger had had in the past six months and the rationale was (Cheng & Holyoak, 1985; pg 401):

> "This is to ensure that entering passengers are protected against the disease".

As well as presenting different rationales, Cheng and Holyoak (1985) also varied the participants' prior experience with the rule. This was done by using different sets of participants, one in America, who were not familiar with the postal rule, and one in Hong Kong, who were familiar with the postal rule. It was expected that few of the participants would have had experience of the cholera rule. Cheng and Holyoak predicted (in line with their pragmatic reasoning schemas approach) that if the rationale evokes a permission schema, then overall performance should be better in the rationale than the no-rationale condition. The only no-rationale condition that would do well would be the Chinese group on the postal rule as they had recent experience of a similar rule. Performance in the conditions was exactly as predicted so that rationales designed to evoke a permission schema facilitated performance on tasks for which participants lacked specific experience.

Cheng and Holyoak's (1985) pragmatic reasoning schemas theory seems to provide strong evidence that reasoning on the selection task improves under a deontic context. Indeed, Cheng, Holyoak, Nisbett, and Oliver (1986) even went as far as to demonstrate that people could do well on abstract rule forms such as '*if action A is to be taken then precondition P must be satisfied'* — but only when a pragmatic reasoning schema was activated through appropriate instructional cueing. Despite the compelling nature of the pragmatic reasoning schemas account, some evidence has raised serious questions about its plausibility. For example, studies have shown that some deontic rules simply do *not* facilitate good performance on selections tasks. Consider the conditional *'if someone stays overnight in the cabin then they must bring a bundle of wood'*. People reason well with this rule only when they are given additional information to encourage them to think about the violating case, that is, someone who stays overnight and does not bring a bundle of wood (Gigerenzer & Hug, 1992).

Another significant theory that attempted to explain facilitation on the deontic task was put forward by Cosmides (1989). She proposed an evolutionary account that suggested that facilitated performance on deontic selection tasks was due to the operation of innate mechanisms, which she called 'Darwinian algorithms'. These mechanisms are highly content-specific and are embodied within innate mental modules. Cosmides argued that we have evolved these mechanisms in order to maximise our ability to achieve our goals in social situations. She was particularly interested in social exchange, where two people must co-operate for mutual benefit, that is, where people make a social contract. This idea of social exchange can be seen in the following rule: '*If you take a benefit, then you pay the cost*'.

Cosmides (1989) proposed that we have developed an innate module that deals with cheater detection and that it has been crucial to do so from an evolutionary perspective. A 'cheater' is someone who breaks a social contract rule: someone who takes the benefit without paying the cost. Cosmides suggested that the facilitation typically observed on deontic tasks occurs because the rules are embedded in these social contracts and so we are able to detect 'cheaters', for example, someone who does not use a high enough value of stamp, or someone who is not 18 years of age.

Cosmides (1989) claimed that a test of her theory would be to switch the social exchange rule to: '*If you pay a cost, then you take a benefit*'. On the normal social exchange rule, when detecting a cheater, participants would chose the *p* and *not-q* cards (also the logically correct cards). On the 'switched' social exchange rule, however, participants would still be looking for the cheater but this would now correspond to the *not-p* and *q* cards (the reverse of the logically correct cards). Cosmides' prediction was supported.

One key problem with Cosmides' theoretical position however, is that facilitation in the selection task occurs for conditionals that are *not* social contracts, such as precautions (e.g., Cheng & Holyoak, 1989; Girotto, Blaye, & Farioli, 1989; Hiraishi

& Hasegawa, 2001; Manktelow & Over, 1991). Even children as young as nine years of age seem to be able to reason very effectively about avoiding dangerous risks (Girotto et al., 1989). Such findings have led Darwinian-algorithm theorists to add an innate module for hazard management (see Cosmides & Tooby, 1992). As Byrne (2005) points out, however, there are other kinds of deontic rule that seem distinct from social contracts, costs and benefits, or hazard management, such as: *'if we are to take care of the planet then we must plant more trees'* or *'if you want spiritual enlightenment then you must meditate'*. Domain-specific accounts that are based on Darwinian algorithms or pragmatic reasoning schemas may well just be far too restricted in their scope of explanation (Byrne, 2005).

Even more damaging for these domain-specific views is recent evidence that not all social-contract conditionals will facilitate appropriate card selections (e.g., Liberman & Klahr, 1996; Love & Kessler, 1995). As such, it does not seem to be the case that a social-contract situation is sufficient to facilitate conditional reasoning. In addition to social-contract rule-content, it seems crucial that participants are cued directly with an explicit 'cheater-detection' framework (Gigerenzer & Hug, 1992). However, asking people to check for cheaters — rather than to check whether a social contract has been followed — seems to transform the selection task into a trivially straightforward categorization task (see Sperber & Girotto, 2002, 2003). Thus, the 'check-for-cheaters' instruction cues people simply to indicate which presented cards would be an example of the category (i.e., cheater) that is predefined in terms of a specific combination of negative and positive traits. People tend to be readily capable of identifying instances of a category when they know (or are told) its characteristic traits (Sperber & Girotto, 2002) regardless of whether the category has evolutionary significance (like a person who is a cheater) or not (like a glider).

1.5.4 Changes so that participants can envisage alternatives

Having had a brief foray into some of the main phenomena associated with thematic—and specifically deontic—versions of the selection task, we now return in this section to consider some key factors that appear to have marked effects on people's success with the selection task when manipulated in the context of the standard abstract and indicative form of the conditional rule.

One important task variant that was seen to have a profound facilitatory effect on card selections was the so-called 'reduced array selection task' or RAST that was first introduced by Johnson-Laird and Wason (1970). The RAST was specifically designed to encourage participants to think only about consequent choices on the selection task, with the idea that logical performance might be seen to increase. In the original Johnson-Laird and Wason (1970) study, participants were asked to verify or falsify a rule statement such as '*If they are triangles then they are black*' that described the contents of two boxes. In order to perform this task, participants were required to ask for objects either from a box that contained only black figures or from a box that contained only white figures. According to logic, they clearly need to search the white box exhaustively and they do not need to look in the black box at all. All participants eventually solved the problem and instructions that asked the participants to falsify the rule promoted better performance than instructions that asked them to verify the rule.

Subsequent research using the RAST (e.g., Wason & Green, 1984; Roth, 1979) has provided strong evidence for facilitated logical responding using the standard selection task but with the antecedent (*p* and *not-p*) choices absent, or even with the *p* card replaced with another *not-p* card. Although these RAST manipulations increase the selection of the *not-q* card, it has also been shown that when the standard four-card task is presented immediately after the RAST there is no transfer of the facilitation effect (Wason & Green, 1984).

Another way to manipulate the abstract, indicative selection task so as to facilitate logical responding involves altering the wording of the associated instructions. Yachanin and Tweney (1982) explored this factor in relation to abstract versus thematic tasks in order to test the idea that the thematic-facilitation effect might have something to do with the difference in the wording of the instructions on the thematic task relative to the abstract one. They noted that abstract tasks typically use the 'true-false' instruction, for example requesting the participant to determine *'which card or cards need to be turned over in order to decide whether the rule is true or false'*. Yachanin and Tweney argued that this instruction may be more challenging for reasoners than asking them to decide whether or not the rule has been *violated*, as is the typical instructional request with thematic versions of the task. On an abstract version of the task Yachanin (1983, Experiment 1) did indeed find that performance could be facilitated using a violation form of instruction, suggesting some support for the hypothesis that a violation set may help cue the identification of falsifying cases.

1.6 Conditional Reasoning Theories as Applied to the Selection Task

We have now seen that there is a vast wealth of research that has been conducted on both abstract and thematic selection tasks over the 40 years since Wason's original formulation of the standard indicative version of the four-card problem. In fact, the previous review has been highly selective, aiming more to give a flavour of some of the key studies and findings relating to the selection task rather than to provide a full overview of the many hundreds of studies that have now been pursued using the paradigm. As such, the review reflects only a relatively small proportion of the vast amount of work that has been conducted with the selection task. Although we have dealt with some important theoretical concepts—specifically the notions of pragmatic reasoning schemas and Darwinian algorithms—we now progress toward a far more in-depth consideration of theoretical issues in the present section.

The following review of conceptual issues will focus almost exclusively on theories that have attempted to account for the pattern of performance observed on the abstract, indicative selection task, as this task is the main focus of the present thesis. As a reminder, participants make two specific errors on the affirmative *'if p then q'* version of the abstract selection task: (1) they incorrectly select the *q* card, and (2) they fail to select the *not-q* card. More generally, when using the full negations paradigm with this task the evidence indicates a general tendency across all logical cases for participants to select more *matching* cards (i.e., cards named in the conditional rule) than *mismatching* ones.

Over the last 20 years there have, in fact, been significant advances in explanations of conditional reasoning with the abstract selection task, with three main accounts tending to dominate theorising: (1) the mental models theory of Johnson-Laird and Byrne (e.g., 1991, 2002); (2) the heuristic-analytic theory of Evans (e.g., Evans, 1984, 1996, 2006; Evans & Over, 1996); and (3) the information gain theory of Oaksford and Chater (e.g., 1994, 1995, 2001, 2003). All of these theories are, in fact, aimed more broadly at accounting for conditional reasoning beyond the selection task (and, indeed, reasoning beyond the conditional rule form). For the purpose of the present thesis, however, the primary focus will be on how these three theoretical accounts have lent themselves to detailed explanations of response patterns that are observed on the abstract selection task.

To facilitate this theoretical review, the following discussion will take each theory in turn, and first present a summary of the overarching assumptions of the theory, before then progressing to a critical discussion of how the theory may account for abstract selection-task performance. The way in which each of these theories have been applied to selection tasks involving content-based conditionals will mostly (though not entirely) be left aside in an attempt to limit the scope of the review to the abstract task that is so central to this thesis. This is despite the fact that each of these three theories has been extended so as to afford detailed, domain-general accounts of thematic selection tasks that avoid the pitfalls of the domain-specific models critiques earlier (i.e., pragmatic reasoning schema theory and the Darwinian algorithm account). For extensions of these theories to thematic selection tasks the reader is referred to Byrne (2005) in the case of mental models theory, Evans and Over (1996) for the Heuristic-Analytic theory, and Oaksford and Chater (1994) for the information gain theory.

1.6.1 Mental models theory and the selection task

The early versions of the mental models theory that were proposed by Philip Johnson-Laird (e.g., 1983) were particularly well-developed in relation to syllogistic reasoning. The basic theory, however, has since been the subject of many revisions and reformulations and has been extended to propositional, inductive and probabilistic reasoning (e.g., Johnson-Laird & Byrne, 1991, 2002; Johnson-Laird, 1995, 2001; Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 1999) as well as counterfactual thinking (e.g., Byrne, 2005). The theory proposes that the mind can be conceptualised as embodying a set of procedures that are used to manipulate internal representation in the form of 'mental models' (cf. Craik, 1943, and Wittgenstein, 1953, for important forerunners of this idea). Mental models are analogue representations of how the world would be if the premises of an argument were true. As such, the theory proposes that reasoning is *semantic* in nature and, thereby, depends upon semantic procedures for constructing and evaluating mental models.

As part of its procedural semantics the mental models theory encompasses three reasoning stages (Johnson-Laird & Byrne, 1991). First, the reasoner has to interpret and understand the presented information such as the premises of a syllogistic argument. They do this by *constructing* models of the possible states of affairs conveyed by the premises. The second stage is the *description* stage where the reasoner has to combine the premises to obtain a description of the state of affairs that they jointly represent. This description has to include something that the first model did not represent, or was not previously explicit in a premise, for example, a putative and semantically informative conclusion in the case of a syllogistic reasoning problem. If the reasoner is unable to produce a description of this nature then no conclusion is produced at this stage. If the reasoner can produce such a description they move on to the third stage which is the *validation* process. This involves the reasoner searching for alternative models that are consistent with the premises but where the putative conclusion is false. If the reasoner finds these alternative models then the conclusion is false and another conclusion needs to be searched for and validated until there are none left. A conclusion is only valid if there are no alternative models that falsify it.

The theory uses a formal notation to depict the mental representation of each premise and the conclusion (e.g., see Johnson-Laird & Byrne, 1991). This notation can be demonstrated by examining at how the model theory accounts for the relative ease with which people are known to draw the valid Modus Ponens inference in conditional reasoning compared with the relative difficulty that they have in drawing the valid Modus Tollens inference (Johnson-Laird & Byrne, 1991). Consider the following major premise:

If there is a circle then it is red

and the subsequent minor premise:

There is a circle

The notation associated with the theory would lead to the representation of the major premise as follows:

[circle]   red

...

Different models are represented on different lines with the square bracket formalism indicating that circles are exhaustively represented with respect to red. This means that circles cannot occur in any other model unless red also occurs in that model. The three dots (ellipses) underneath this model denote the existence of implicit models (i.e., further models of the rule that are possible but which have not yet been 'fleshed out' and made explicit).

When given the additional information 'there is a circle' as a minor premise, it follows that since 'circle' is exhaustively represented in the explicit model, then 'it must be red'. The ease of deriving this inference from the minor premise corresponds with the typically high rate of Modus Ponens inferences observed in the conditional inference paradigm.

If we turn to the Modus Tollens problem, the major and minor premises would be as follows:

If there is a circle then it is red
It is not red

In this case no inference seems to follow immediately from the model of the major premise:

[circle] red

This is because the case 'not-red' (i.e., ¬red where the symbol '¬' denotes negation) is not explicitly represented. In order to draw the Modus Tollens inference, Johnson-Laird and Byrne claim that the model needs to be fleshed out to represent all of the true possibilities, which would require the reasoner to add extra models to their mental representation as follows:

[circle] red
¬circle ¬red
¬circle red

The reasoner then has to grasp that the minor premise 'it is not red' rules out the first and the third model, and because of this the conclusion follows that 'there is not a circle'. The inference is made less often because not all reasoners will be successful at pursuing the fleshing out of additional models.

Johnson-Laird and Byrne (1991, 2002) have also included an important psychological constraint in the theory known as the 'principle of economy'. This suggests that due to restricted working memory capacity reasoners will do as little work as possible, that is, they will construct the minimum number of models with the minimum amount of material represented explicitly. This explains why reasoners may fail to construct counterexamples to current models or why they may be heavily influenced by the effect of content because prior knowledge adds information to the models that are constructed.

So how does mental models theory deploy its core assumption to provide an account of the response patterns observed for the standard abstract version of the selection task? Let us start by examining the situation for an affirmative conditional rule such as '*if there is an A on one side of the card then there is a 3 on the other side*'. This rule would be represented as:

[A] 3

...

The square brackets around the 'A' indicate that the A is represented exhaustively in relation to the 3 (i.e., whenever an A occurs a 3 also occurs). The ellipses denote an implicit model (i.e., that there are further models of the rule that are possible that have not yet been made explicit). According to the principle of economy the reasoner will construct the minimum number of models with the minimum amount of material represented explicitly. In other words they will tend to construct only the model as depicted above. To account for selection task performance, however, the mental models theory makes two further assumptions. First, participants only consider those cards that are explicitly represented in their models of the rule (i.e., the 'A' and the '3' cards in this case). Second, participants only go on to select those cards for which the hidden value on the reverse side of a card has a bearing on the truth or falsity of the rule (i.e., the 'A' card). Thus the failure to select the falsifying 7 card (i.e. *not-q*) on the '*if A then 3*' rule reflects the fact that this term is not explicitly represented in the reasoner's models of the rule.

The mental models account additionally proposes that some people will represent the rule as a biconditional (i.e., with '*if A then 3*' implying its converse '*If 3 then A*'), as follows:

[A] [3]

...

People who represent the rule as a biconditional would select both the 'A' card and the '3' card (the other popular selection combination that is observed on the task), as both these cases are explicitly represented in models and could bear on the rule's truth or falsity.

In order to account for matching bias on the selection task with rules that contain negations, Johnson-Laird and Byrne (1991) originally suggested that negated components give rise to the expansion of models to include the affirmative counterparts of negated terms. Such affirmative (matching) values would then lead to card selections if hidden values impacted on the rule's truth of falsity. For example, a negated-consequent rule such as *'if A then not 3'* would be represented as follows:

[A] ¬3
 3
...

Since both the 'A' card and the '3' card have hidden values that impact on the truth or falsity of the rule, and since these values are both explicitly represented in models, then these cards would tend to be selected by participants. This is precisely what is seen in selection-task studies (e.g., Evans, 1984).

In their more recent formulation of the mental models theory, however, Johnson-Laird and Byrne (2002) have discarded the suggestion that the affirmative counterparts of negated terms are added to mental models. Instead, they now subscribe to Evans and Handley's (1999) mental model theory revisions that have overcome contradictions in its account of the selection task when viewed in conjunction with the account of conditional inference tasks. This new mental models theory of the selection task embodies the 'simple unadorned principle of truth [whereby] mental models represent true assertions, whether they are affirmative or negative, but not false assertions' (Johnson-Laird & Byrne, 2002, p.699). According to this view, all conditionals are initially modelled with only the True Antecedent and True Consequent cases. Reasoners attempt to combine *each* card value with the model to draw an inference about the logical consequence of hidden values. Matching bias arises in the course of *fleshing out* models. For example, in trying to combine the critical False Consequent card (e.g., 7) with the model of the rule *'If A then 3'*, nothing appears to follow (see below) because of its implicitly-negated (i.e., 'non-matching') status:

[A] 3
    7

With these recent revisions, then, the mental models theory can not only provide a generic account of how conditional rules are represented across different conditional-reasoning tasks that are studied in the literature (including the selection task and conditional inference tasks), but it can also afford a compelling explanation of the pattern of matching card selections seen on the indicative selection task using the full negations paradigm.

1.6.2 Heuristic-analytic theory and the selection task

Evans (1984, 1989) proposed a theory of reasoning, known as the heuristic-analytic theory. This was later developed and extended by Evans and Over (1996) into what they referred to as 'relevance theory', although recently Evans (2006) has reverted to the 'heuristic-analytic' distinction in a revised version of his earlier heuristic-analytic theory. Evans' new heuristic-analytic theory fully embraces ideas about dual processes determining reasoning responses. The heuristic-analytic theory did, in fact, develop out of an earlier dual-process theory first proposed by Wason and Evans (1975) in an attempt to account for their observation that the verbal explanations that participants gave for their card choices were discrepant from their actual card choices. Wason and Evans (1975) tested participants using two rules implementing the use of negatives as proposed by Evans and Lynch (1973). Participants received conditional statements of the form *'if A then 3'* (abstract affirmative) first and then *'if A then not 3'* (abstract negative) and were asked to give verbal justifications for their choices on the task. Results showed that regardless of the existence of a negative in the rule, card selections remained the same (i.e., choices still reflected the matching bias effect with participants choosing both the p and q cards). On the affirmative version of the rule this choice is incorrect according to normative logic, whilst on the negative version of the rule this choice is correct according to normative logic. However, although card choices remained the same, verbal justifications were found to be very different. When given the negative rule participants would provide justifications in terms of falsifications, so they would talk about disproving the rule, whereas on the affirmative rule participants talked about proving the rule. Wason and Evans' (1975) dual-process theory attempted to account for these observations by suggesting that card selections were caused by an unconscious matching bias and that the verbal explanations made were really just rationalisations of choices.

The heuristic-analytic theory of Evans that was proposed through the 1980s and 1990s and up to the present day (e.g., Evans, 1984, 1989, 1996, 2006) provides a re-formulation of the earlier Wason and Evans (1975) dual-process account. In essence, the heuristic-analytic theory proposes that it is first important to determine what people are reasoning *about* before it is possible to make theoretical claims about *how* they are reasoning. In all of his heuristic-analytic theories Evans proposes that reasoning takes place in two stages. First there is a *heuristic* stage where pre-attentive processes deem certain information as relevant and select it for further processing. Second, there is an *analytic* stage which serves to generate an inference or judgement from the selected information. This is done consciously using explicit processes. Evans argues that any information that is deemed to be irrelevant at the first, heuristic stage is unlikely to be processed further. Evans (e.g., 1989) has further argued that heuristic processes are *unconscious* because of their pre-attentive nature, that is, heuristic processes primarily function to determine *what* participants will attend to and think about (i.e., what they will see as being 'relevant' to the task at hand). Errors and biases can, therefore, occur during reasoning either because logically relevant information is selected out or because logically irrelevant material is selected in at the heuristic stage. Evans' heuristic-analytic theory is, therefore, first and foremost an *attentional* one.

Evans' heuristic-analytic theory (e.g., Evans & Over, 1996) is closely linked to other dual-process theories of cognition as discussed above in relation to the rationality debate, including Reber's (1993) dual-process account of implicit and explicit learning, and Stanovich's (1999, 2004) dual-system framework. According to the dual process account of reasoning, both System 1 and 2 cognition are present in all human beings (Evans, 2003). The System 2 in Reber's theory is the part that is linked to Evans' (1989) analytical stage of reasoning. This system provides humans with many advantages over animals. It is the part of the system that allows us to represent possibilities (as shown across a variety of different reasoning and judgement tasks; e.g. see Evans and Over; 1996). As with higher animals, we can make decisions based upon our past previous experience, which is often shaped by reinforcement. This is System 1 decision making. However, we can make decisions differently to this if we want to, as we can think about future consequences and imagine possible outcomes of our decisions. We may make these decisions after analysing the probability and utility of these possibilities. This is System 2 thought. This ability to engage in System 2 thinking is the thing that distinguishes us from other animals as it is not available to them. The psychological literature, however, is full of evidence of biases and failures to make consequential decisions or provide logical errors and occurs because most of the time people are engaging in System 1 thinking.

If we apply the heuristic-analytic theory to the abstract selection task, Evans argues that evidence for the effect of *relevance* is reflected in biases such as matching, whereby cards appear relevant when their features match those named in the conditional rule. Within the heuristic-analytic theory the biases that are claimed to determine card selections are, in fact, attributed to the operation of two heuristics: the so-called *if-heuristic* and the *not*-heuristic (the latter is also referred to as the *matching* heuristic). The suggestion is that the linguistic function of 'if' enhances the relevance for True Antecedent cases because it directs attention to the situation where the antecedent is true and away from the situation where it is false. The linguistic function of 'not' is to direct attention to the proposition it denies. Evans argues that negation is not commonly used in natural language to assert new information but, instead, to deny presuppositions (see also Wason, 1966). For example, if we take the sentence 'Today I went shopping' and then add 'not' so that it becomes 'Today I did *not* go shopping', we can see that the *topic* of the sentence remains the same, that is, it is still about shopping, but the *comment* (or meaning of the sentence) has changed. Because the topic remains the same, participants stay focussed on this and still choose cards that are mentioned in the rule (i.e., the presence of negations has little if any effect on card selections in the selection task).

In summary, then, when faced with an abstract task the heuristic-analytic theory proposes that participants rely purely on linguistic cues to guide their card selections (i.e. card selections are driven by the function of the *if* and *not* heuristics). However, when presented with a context, as happens in a realistic version of the rule, selections are claimed to be determined by pragmatic cues (rather than linguistic ones) and participants use content-specific rules of inference. One suggestion is that people's card selections may be guided by specific knowledge of counterexamples to stated rules that derive from their memory for similar situations (Griggs and Cox, 1982). For example, in the Drinking Age problem participants have prior knowledge about drinking laws and cases of rule violations that may be cued directly by the use of a police-officer scenario. In essence, the given rule and scenario make certain information appear relevant, leading to the correct response of the *p* card and the *not-q* card so as to detect a potential violating instance of the rule. Because the pragmatic cues are stronger in these contexts than linguistic cues, participants do not succumb to matching bias on realistic versions of the selection task.

Evidence for the heuristic-analytic theory in terms of matching bias has also been identified on other conditional reasoning tasks. The truth table task, for example, requires that participants indicate which logical cases verify and which falsify conditional statements. Matching bias is displayed when mismatching cases are more likely to be classified as irrelevant. Conversely, relevant cases are sorted into true and false and these decisions are clearly determined by logic. There is also considerable evidence to suggest that matching bias is related to implicit negation on a task. This can be seen on the selection task when the *not-q* card (e.g., a '7' card) is implicitly represented as something that is not a *q* (i.e., that is not a 3). If explicit negation is utilised in the abstract selection task (i.e., with cards showing values such as 'A', 'not-A', '3' and 'not-3') then the matching bias effect should be eliminated because the lexical content of the rule and the card values will always match. This has been demonstrated on the truth table task as shown by Evans (1983) and later by Evans, Clibbens, and Rood (1996). Evans et. al. (1996) permuted negation through a conditional rule and found a significant reduction in matching bias and an increase in logically correct responding. The results of this experiment were extended to a second experiment using the selection task. Again, the use of explicit negation resulted in a decrease in matching bias. However, there was no increase in logically correct choices (see also Griggs & Cox, 1993; Jackson & Griggs, 1990; Kroger, Cheng, & Holyoak, 1993). One important issue here concerns the difference in effects of explicit negation in terms of logical responding that is seen across the two paradigms that both involve the same conditional rules. Why is it that on the selection task matching bias can be decreased with the use of explicit negation on cards, but there is no change in logical responding on the task? Evans et. al. (1996) argues that it is because on the selection task participants are asked only to indicate the cards they *would* choose, and so the task may not actually get people involved in explicit reasoning. The task does nothing more than ask participants which cards are *relevant*, and it therefore only invokes the use of the heuristic part of the system. In contrast, the truth table task requires not only a judgement of relevance but also a further judgement of whether the cases are true or false and so inevitably requires an analytical reasoning stage.

This is an important difference between the two tasks, and in light of this finding Evans and Over (1996) revised certain aspects of the heuristic-analytic theory. One important revision concerned the specification of the processes involved in each stage. In the heuristic-analytic theories that pre-dated Evans and Over (1996), heuristic processes were only defined as preconscious processes which produced an explicit representation of the relevant information. There then had to be some sort of analytical processing to produce an observable outcome. It was assumed, then, that the analytic stage took over from where the heuristic stage left off. Evans and Over (1996) instead proposed that although tacit processes are responsible for relevance and focusing, they may also lead directly to judgements and actions. Therefore, the theory proposed that explicit reasoning processes *may* determine decisions, *but need not do so*. Accordingly, on conditional truth table tasks explicit processes *do* lead to correct logical choices, as we have seen, but in the case of the selection task Evans and Over (1996) claimed that participants' card choices are directly led by implicit, heuristic processes (i.e., card choices are purely reflected by relevance). Indeed, Evans and Over argued that participants do not engage in an analysis of the logical consequence of turning each card but that time spent thinking on the selection task is used to *rationalise* the choice that immediately appears relevant. Explicit reasoning, then, may well occur on the selection task but it does not affect card choices. It is this latter point that critically distinguishes the heuristic-analytic theory from the mental models theory–which assumes that analytic processing not only arises on the selection task but is functional in determining card selections (i.e., cards are only selected if they impact on the rule's truth or falsity). Evans' (e.g., 1989) view is that any analytic processing on the task serves *only* to rationalise heuristically-determined choices.

Some obvious questions that seem to arise at this point in relation to the heuristic-analytic theory of reasoning might be 'what is the purpose of relevance?' or 'why is it useful to limit the information that passes through to the explicit, analytic system?' Evans and Over (1996) argue that in order to function intelligently we need to be able to select from the huge amount of information that is available to us from our memory and from our environment. The preconscious heuristic stage therefore selects information that is perceived to be relevant. Conscious thought is applied to this selected information and this is the most efficient and effective way for us to reason and survive an environment saturated with complex information.

Recently, Evans and Over (2004) have embodied the heuristic-analytic theory within their broader 'suppositional' theory of conditional reasoning, whereby conditionals focus attention on the supposition that the antecedent case hold true (a view entirely consistent with the *if heuristic*) such that people will develop a single 'epistemic mental model' that, by default, incorporates the True Antecedent case. More recently still, Evans (2006) has linked his ideas about heuristic and analytic processes (as well as his notion of 'suppositional' thinking) to a far broader and encompassing dual-process concept of reasoning that incorporates what he refers to as 'three principles of hypothetical thinking'. The principles are: the *singularity principle*, the *relevance* *principle*, and the *satisficing principle* (see also Evans, Over, & Handley, 2003). Evans (2006) invokes these principles to argue that people will construct only one mental model at a time with which to represent a hypothetical situation (singularity principle). This model is pragmatically (i.e., heuristically) cued to be the most relevant within the prevailing context (relevance principle). By default, this model will represent the most probable or believable state of affairs. Finally, this model will be subject to explicit (i.e., analytic) evaluation that complies with a satisficing principle.

Evans (2006) argues that the singularity principle derives from the consideration that the analytic system, whilst at the core of hypothetical thinking, has severely limited processing capacity. Evans proposes that the other two principles reflect the functioning of the heuristic and the analytic system. The relevance principle is embedded within the heuristic system, whose primary purpose is to deliver both knowledge-based and belief-based content to the analytic system. The satisficing principle, on the other hand, is taken by Evans to reflect a fundamental bias in the analytic system to operate on representations that it has been provided with via the heuristic system, unless there are good reasons to 'give up' or overturn these representations. Such satisficing-oriented processing, however, is deemed to be perfectly rational in a world where one typically has to make decisions relatively quickly without an endless analysis of possibilities (cf. Simon, 1982), who originally coined the notion of satisficing as a boundedly-rational approach to decision making in situations where it is impossible to optimise).

The revised heuristic-analytic theory can continue to explain many cognitive biases in reasoning in a similar way to the original theory of Evans (e.g., 1984, 1989). That is, biases can arise because the heuristic system fails to represent logical features of the problem as being relevant or because it represents features that are logically irrelevant to the task at hand. Evans (2006) suggests that the evidence indicates that such heuristically-generated biases can be inhibited—at least to some extent—by the operation of the analytic system, which intervenes in order to 'reset' default epistemic mental models delivered to it by heuristic processing. However, the analytic system is, itself, viewed as being prone to biases of its own, particularly those that arise from the operation of the satisficing principle (i.e., the tendency for analytic processes to hold on to representations that are merely 'good enough', which leads to the frequently observed endorsement of fallacious inferences as well as confirmation tendencies in hypothesis testing).

Evans' (2006) heuristic-analytic account of the abstract selection task is really no different to his prior accounts, except, perhaps, in the fuller recognition that for most participants the analytic system *is* seen to be actively engaged in the task (which contrasts somewhat with Evans et al.'s, 1996, view – but not with the more developed position of Evans & Over's, 1996). At the same time, the argument remains that any analytic processing that does occur on the selection task does not serve to alter default heuristic responses, but instead functions merely to *rationalise* such responses. The critical new element of the revised heuristic-analytic account of the selection task is the proposal that the satisficing principle can provide a neat account of why the analytic system fails to override heuristically cued card choices. In particular, Evans (2006) suggests that most people (except those of very high cognitive ability) will treat verification and falsification on the task as though they are *symmetrical*. Thus, participants will happily justify a choice of a matching card combination on the grounds that it will prove the rule true or prove it false (as was, in fact, originally shown by Wason & Evans, 1975). Since the standard task instruction refers to discovering 'whether the rule statement is true or false', Evans argues that the analytic system simply satisfices (accepting the heuristically-cued choice) whenever it can find a verification *or* falsification justification. In practice, of course, this means that heuristically-cued choices on the selection task will nearly always be accepted by participants.

Recently, there have been a number of studies that propose that analytic reasoning *does* alter participants' card selections on the abstract selection task. In particular, studies by Feeney and Handley (2000) and Handley, Feeney, and Harper, (2002) have presented a collection of experiments that present participants with a second conditional rule that expresses an *alternative* antecedent to that contained within the

first conditional rule. For example, Feeney and Handley (2000) presented participants with an affirmative, indicative selection task with the rule:
'If the card has a letter A on one side then it has the number 3 on the other side'.

A second group were also presented with the rule:

'If the card has a letter L on one side then it as the number 3 on the other side'.

The outcome of this additional rule was a *decreased* tendency for people to select the *q* card. Feeney and Handley explain this finding by suggesting that participants actually are considering the unseen side of the cards and recognising the fact that there might be either *p* or a letter that is not a *p* on the other side of the card. Although this finding is at odds with the heuristic-analytic account, it can be argued that the manipulation that Feeney and Handley introduce is significantly altering the nature of the task such that the alternative antecedent possibility puts into question the original rule, thus, allowing participants to think more thoroughly through the consequence of turning the *q* card.

Other work by Roberts and Newton (2001) also makes a claim for analytic processes playing an important part in selection-task performance. In two of their experiments, they presented participants with a rapid response selection task where participants had to respond to each card within 2 seconds of its presentation. They found that when time was curtailed in this manner matching responding increased slightly compared to that on the free-responding version of the task. Roberts and Newton propose that with available time—as on the free-responding condition—analytic processes are responsible for the overturning of some matching responses. Roberts and Newton did not, however, find statistically significant evidence for markedly improved logical performance on the free-time version of the task relative to the rapid-response version. The findings and implications of these studies will be considered in more detail later on in the thesis.

1.6.3 Information gain theory and the selection task

The concept of 'information gain' comes from the technique of 'rational analysis' developed by J.R. Anderson (e.g., 1990, 1991), and was applied to the selection task by Oaksford and Chater (e.g., 1994, 1996; see also Oaksford & Chater, 2003). The information gain theory is a general theory of reasoning and takes the view that cognitive behaviour is adapted to the structure of the environment and does not depend on rules or models that other theories of reasoning propose. It is essentially a theory at the *computational level* (i.e., a theory that attempts to state what the mind is computing). It proposes that when in a reasoning experiment, participants are not trying to test whether a rule is true or false, for example, but are instead looking for information that will help them update their beliefs. Oaksford and Chater propose that when engaged in reasoning our main aim is to gain information as this leads to a decrease in uncertainty. The information we select in order to help us reduce uncertainty is known as *optimal data.* They suggest that card selections in the selection task are based on the information value of each card estimated in the form of *expected information gain.* Oaksford and Chater's mathematical analysis of the information value of cards shows, for example, that the selection of the matching *q* card for the affirmative conditional can be more useful than the selection of the non-matching (but logically appropriate) *not-q* card. In this way, the information gain theory of the selection task proposes that illogical matching choices may, in fact, be deemed to be *rational* in terms of a probabilistic standard. It is important to note that Oaksford and Chater's (e.g., 1994, 1995) analysis of the selection task also entails a *rarity assumption,* which is that most properties of the world (including the properties described by *p* and *q* in selection task studies) apply to a small set of objects, and that people's strategies for testing or framing hypotheses are, by default, adapted to situations where rarity holds (for supporting evidence see Anderson & Sheu, 1995; McKenzie, Ferreira, Mikkelsen, McDermott, & Skrable, 2001; McKenzie & Mikkelsen, 2000).

The information gain theory presents a persuasive account of the matching effects observed on affirmative conditional rules within the selection task. Moreover, because the theory capitalises on Oaksford and Stenning's (1992) arguments that negations typically define high-probability contrast sets, it is also readily able to explain antecedent and consequent matching effects observed for conditional rules containing *negated* constituents (e.g., Oaksford, 2002a; Yama, 2001). So, for example, a rule such as '*If there is an A on one side of the card then there is not a 3 on the other side*' is argued to designate a high probability true consequent category (any number that is *not a 3*), whereas the false consequent category is represented by a very low probability single case (the matching '3' card), whose rarity assures its high information value.

Overall, then, the information gain account is able to accommodate a wide range of evidence for matching effects in the standard selection-task paradigm. A final strength of the theory—and one which sets it apart from both the heuristic-analytic and the mental models accounts—is its capacity to explain the considerable body of evidence that has now been amassed for probabilistic influences on card selections (e.g., Green & Over, 1997; Green, Over, & Pyne, 1997; Kirby, 1994; Oaksford, Chater, & Grainger, 1999; Oaksford, Chater, Grainger & Larkin, 1997). So, for example, it has been shown that card selections vary in ways predicted by information gain when *P(p)* and *P(q)* are varied experimentally. Non-probabilistic theories are generally not readily able to explain why probability manipulations should affect card selections, only really being able to do so by invoking the idea that participants adopt different task interpretations, with probabilistic manipulations affecting the proportion of people adopting these different interpretations (see Oaksford & Wakefield, 2003).

In spite of the capacity of the information gain theory to explain an impressive range of selection-task data, it has been claimed to have certain limitations. One problem (cf. Evans, 2002) is the difficulty that the theory appears to have in explaining why the use of *explicit* negations on cards in selection tasks completely removes matching bias (e.g., Evans et al., 1996). This phenomenon is easily accounted for by the heuristic-analytic theory, as all cards present matching values within an explicit negations paradigm. Oaksford (2002a), however, has recently proposed that this explicit negations effect may be a result of participants failing to engage their 'normal' interpretative processes in this task variant—an explanation that is certainly worthy of further investigation.

1.7 Conclusions to Chapter 1

This chapter has provided an introduction to the psychological study of human reasoning with conditional statements. A review of some of this work has shown how central the issue of rationality is in understanding people's behaviour on conditional reasoning tasks. In particular, the chapter has focussed on the range of findings observed on the Wason selection task, where studies have embarked on a number of manipulations (e.g. permuting negatives, the introduction of thematic and deontic content, and altering the structure of the task, such as the number of cards or the instructions given) in an attempt to improve logical responding on the task.

Another main driver in the selection-task literature has been to explain the common pattern of logical errors observed on the abstract, indicative task, in particular the evidence for responding being subject to a *matching bias*. In this respect, the present chapter provided a detailed introduction to three contemporary theories of the selection task that have all dedicated considerable effort to an explanation of both matching bias and other reported selection-task phenomena. The mental models theory (e.g., Johnson-Laird & Byrne, 1991, 2002), the heuristic-analytic theory (e.g., Evans, 1984; 1989; 2006; Evans & Over, 1996) and the information gain theory (e.g., 1994, 1996; Oaksford & Chater, 2003) have all been shown to provide compelling accounts of selection-task responses, in particular the matching-bias effect.

To distinguish between these aforementioned theoretical positions, however, an approach is needed that allows the researcher to go beyond the actual selection patterns that arise during selection-task performance. Such an approach can, it is argued, take the form of methodologies that enable the collection of process-tracing data. In particular, it is proposed that process-tracing techniques—including eye-movement tracking and think-aloud reporting—can uncover valuable information concerning *what* people are reasoning about when they are tackling abstract selection tasks. Another process-tracing methodology, mouse-tracking—where participants indicate what they are thinking about using a mouse pointer)—has already been deployed with the selection task, although perhaps with rather limited success for reasons that will be explained in the next chapter. Indeed, because of these limitations mouse tracking was not utilised in the present thesis, whereas eye-movement tracking and think-aloud methods were central to the reported research. The next chapter will review all three methods, think-aloud reporting, mouse-tracking and eye-movement tracking, before progressing to a report of the experimental studies that were pursued.

Chapter 2

Experiments 1 to 3

Chapter 1 presented an overview of three dominant theoretical accounts of people's response patterns on abstract, indicative versions of the Wason selection task: (1) the mental models theory of Johnson-Laird and Byrne (e.g., 1991, 2002); (2) the heuristic-analytic theory of Evans (e.g., Evans, 1984; 1996, 2006); and (3) the information gain theory of Oaksford and Chater (e.g., 1994, 2003). Chapter 1 concluded with the view that all three theories seem to be able to provide persuasive explanations of selection-task responses, such that arbitrating between these accounts purely on the basis of participants' response patterns is unlikely to prove fruitful in the short term, even under various manipulations relating, for example, to task format, rule form or problem instructions. Instead, it was suggested that a more productive way to address the adequacy of existing selection-task theories may reside in the deployment of methodologies that can provide behavioural data beyond card selections by tapping into the cognitive processes underlying task performance to give a process-oriented measure of performance.

There are, in fact, a number of studies that have attempted to address explicitly and directly aspects of on-line processing of the selection task using a variety of so-called 'process-tracing' techniques. Examples of such techniques that have been used in selection-task studies include the 'think-aloud' verbal-protocol methodology, whereby reasoners concurrently *verbalise* all of their thoughts (e.g., Evans, 1995), and the mouse-tracking methodology, which entails participants having to indicate those aspects of the task that they are currently *considering* by means of a mouse pointer (e.g. Evans, 1996; Roberts, 1998b).

The present chapter begins with a review of process-tracing studies that have investigated selection-task behaviour. This review will critically discuss the findings that have emerged from this existing body of research, and will also assess the implications of process-tracing findings for the three selection-task theories presented in Chapter 1. One upshot of this review will be the proposal that eye-movement tracking—which involves monitoring the moment-by-moment attentional transitions that arise during task performance as reflected in the pattern and duration of eye-fixations—may well provide an alternative and potentially more valid process-tracing technique for examining reasoning processes on the selection task. This argument will then pave the way to an overview of the first three experiments in this thesis, which all attempted to deploy eye-movement monitoring to study behaviour on the abstract selection task with the aim of furthering a theoretical understanding of processing on this reasoning problem.

2.1 Verbal Protocols and the Selection Task

Throughout the history of psychological experimentation researchers have been keen to gain an understanding of participants' thought processes and experiences as they carry out tasks using a technique referred to historically as *introspection*. However, the use of introspective methods in cognitive psychology has always been contentious, as it has been commonly assumed that high-level thought processes such as reasoning are not ones that can be introspected upon. For example, Nisbett and Wilson (1977) set out a controversial attack on the use of introspection in social judgement and decision-making research. They essentially argued that people do not have awareness of the processes that determine their behaviour, and that verbal reports can, therefore, be both inaccurate and misleading. Nisbett and Wilson reported a number of experiments on attitude change where participants' reports indeed seemed to demonstrate no apparent awareness of the stimuli influencing attitudes.

More recently, however, Ericsson and Simon (e.g., 1980, 1984) have presented a much more positive view of verbal reports of task-based processing as a form of psychological data, and argue that such verbal reports—or what they term *verbal protocols*—appear to be a potentially invaluable source of information when set alongside standard outcome measures that are derived from other observational methods (e.g., response times and solution rates). Indeed, Ericsson and Simon (1980, 1984, 1993) went as far as to justify their proposals by developing a detailed information-processing theory of cognition within which they locate the actual mechanism by which verbal reports are produced. In relation to this model, Ericsson and Simon argue that verbal reports reflect the current contents of *short-term memory* (or what is nowadays more commonly referred to as 'working memory'). These short-term memory contents, in turn, reflect the focus of what the participant is attending to at a given point in time. As such, verbal data are best viewed as being the 'products' of cognitive processes rather than self-generated descriptions of the processes themselves. This means that the *researcher* has the task of inferring underlying processes from the observable verbal products of such processes, whilst the participant producing the report is not in any way required to self-generate a theoretical account of their own thought processes (in direct contrast to the basic expectation with the use of introspective methods).

Critics who argue against the use of verbal-protocol methods in cognitive research (e.g., Nisbett & Wilson, 1977) propose that participants' verbal reports of their thoughts may not necessarily correlate with any observable behaviours that they produce. In other words, it is possible that the information that is retrieved by a participant at the time of the report is different to that which was functionally important during actual performance on the task. This is a very legitimate concern with the validity of verbal reports, as Ericsson and Simon (1993) themselves acknowledge. In order to avoid this problem, however, Ericsson and Simon suggest that it is advisable for researchers only to gain *concurrent reports*, that is 'think aloud' reports that are collected at the time of experimental testing, as opposed to *retrospective reports* that are collected after a task has been done. The belief that concurrent verbal protocols are tapping into cognitive processing seems to be warranted on the basis of the assumption that thought processes are a sequence of states that each contain an end-product of cognitive processes. These informational 'products'—which may be retrieved from long-term memory, directly perceived and recognised, or generated by means of inferential mechanisms—are reasonably stable, and can therefore be verbalised. Any retrieval processes, recognition processes or inferential processes that deliver this information, however, are not open to verbalisation as they essentially arise at an automatic and tacit level outside of conscious awareness.

Another important claim made by Ericsson and Simon (e.g., 1980, 1993) is that when participants are asked to think aloud, the thoughts themselves largely remain unchanged by the added instruction to think aloud. Only if participants are asked to *explain* or *describe* their thoughts does such 'reactivity' (i.e., distortion of the primary task-based process) arise, because additional thoughts and concepts have to be retrieved such that participants have to attend to information that would not normally be needed to perform the task. Ericsson and Simon (e.g., 1993) describe *three different levels of verbalisation* that can arise during verbal reporting, with each level being associated with a greater degree of potential reactivity in relation to 'normal' task-based processing. *Level one* verbalisation involves the articulation of information that is normally heeded in a verbal form anyway (i.e., it is simply the vocalisation of current thoughts). There are no intermediate processes in this case and the articulation will take no special effort. *Level two* verbalisation involves articulating information that is normally heeded in a non-verbal form. This verbalisation involves the description or explanation of these non-verbal thoughts, but does not bring any new information to bear in the process of task-oriented thinking. Because level two verbalisation involves a re-coding process it has been observed to slightly slow down primary task performance whilst not impacting upon the structure of the process (see Ericsson & Simon, 1993, for relevant evidence). *Level three* verbalisation involves explanation of thought processes. This process is not simply reporting what is currently held in short-term memory but requires the participants to link their thoughts to earlier thoughts and information. Level three verbalisation is, therefore, very likely to impact negatively on the nature of normal task-based thinking. Empirical evidence supports this latter conjecture (see Ericsson & Simon, 1993, for a review).

So, in summary, any concerns about the validity and reliability of verbal reports are avoided from Ericsson and Simon's perspective if three conditions prevail: (1) verbal reports are obtained from the participant concurrent with task performance rather than retrospectively; (2) participants are required only to verbalise information that they are heeding rather than to describe non-heeded information or selective aspects of heeded information; and (3) data are analysed by the experimenter and not by the participant. In the context of reasoning research, Evans (e.g., 1989) has concurred with the key proposals of Ericsson and Simon (e.g., 1993) and argues that participants appear to have little difficulty in performing a concurrent think-aloud requirement when tackling reasoning tasks, and the data that can be collected from such reports seem perfectly valid and effective for tracing the locus of attention during cognitive task performance.

Despite such a positive assessment by Evans (1989) of the potential role for verbal-protocol assessments of reasoning performance, a review of the literature suggests that there are, in fact, only a few published studies that have employed this method in the context of research on the Wason selection task. Moreover, those few abstract selection-task studies that have used verbalisation methods appear to have been in conflict with Ericsson and Simon's (1993) recommendation concerning the vital importance of eliciting concurrent rather than retrospective reports from reasoners. For example, a study by Green and Larkin (1995) utilised a *post-hoc* reporting technique, whereby participants had explicitly to provide reasons for their card selections when prompted by the experimenter. It could be argued that this post-hoc approach tells us very little about the on-line focus of participants' moment-by-moment attentional processing as might be gleaned from the use of a concurrent verbalisation method. Another recent study by Stenning and van Lambalgen's (2002) elicited verbalisations from participants tackling selection tasks as part of a Socratic 'tutorial dialogue' between the experimenter and the reasoner. As interesting as this methodology certainly is, the technique may only have a limited bearing on the issue of individual reasoning processes divorced from the dynamics of didactic conversations between students and tutors. Indeed Stenning and van Lambalgen (2002, p. 281) themselves acknowledge that "Engaging subjects in dialogue undoubtedly changes their thoughts, and may even invoke learning. The relation between the reasoning processes evoked by the standard way of conducting the task, and the processes reflected in subsequent dialogues is a relation that remains to be clarified".

As well as the studies reported above, however, there are other abstract selection-task studies that have utilised more reliable concurrent-reporting methods. Beattie and

Baron's (1988, Experiments 2 and 3) study, for example, provided evidence that participants rarely mentioned alternative cards to the ones that they ended up selecting—an effect that Beattie and Baron viewed as supporting the notion of a heuristically-based matching process in line with Evans' (e.g., 1984) heuristic-analytic theory. Participants were also seen to be overconfident about their card choices and showed little sensitivity to the correctness of their selections. However, Beattie and Baron's protocol coding scheme functioned at a fairly gross level of analysis that focused on the classification of selection patterns and the categorisation of responses to probe questions. Because of this, their coding scheme does not appear to have been geared toward uncovering insights into the spontaneous analytic processing that might be associated with card choices.

More recently, Evans (1995, Experiment 5) presented another protocol-analysis study of behaviour on the abstract, indicative selection task. Protocols were analysed in two distinct ways. First, they were scored for references to the *facing sides* of cards. The percentage references were then divided according to whether the participant selected the card or not. Second, protocols were scored for references to the *hidden sides* of cards, and again these scores were broken down according to whether the card was selected or not. Evans' analysis revealed that participants referred more often to the facing sides of the cards that were selected than to the facing sides of cards that ended up being rejected. More importantly, the second analysis showed an identical tendency for participants to refer more to the hidden sides of selected cards than to the hidden sides of non-selected cards. These findings were consistent with Beattie and Baron's (1988) results. Evans argued that the findings also lent further support to his heuristic-analytic theory, which claims that people only think about some of the cards and not others, and that thinking about hidden sides of cards mostly serves to rationalise decisions to choose such cards.

As these previous studies show, think-aloud verbal protocols appear to be a useful methodology for use in studying reasoning processes. Although the protocol method has been applied to a limited extent in studying processing on the abstract selection task, it remains the case that there are no large-scale studies of the task that have collected protocol data.

2.2. Mouse Tracking and the Selection Task

All of the various instantiations of Evans' heuristic-analytic theory (e.g., Evans, 1984, 1996, 2006) are consistent in assuming that selection-task responses for the majority of reasoners can be accounted for in terms of the operation of heuristic processes. Moreover, although some versions of the heuristic-analytic theory (e.g., Evans et al., 1996) did seem to lean toward the proposal that participants do not engage in *any* analytic reasoning on the selection task, Evans (e.g., 2006; see also Evans & Over, 2004) has recently clarified his position by proposing that analytic processes *are* indeed applied on this task, but they only serve to *rationalise* choices (via a satisficing-based mechanism) and play no actual part in altering those choices—except, perhaps, in the case of those of higher intellectual ability, where analytic processing may be sufficiently powerful to overturn default responses that are delivered by the heuristic system.

Fundamentally, then, the heuristic-analytic theory of the selection task claims that card choices are *not* affected by any explicit, analytic reasoning but are, instead, driven primarily by the operation of implicit, heuristic, processes. Evans (e.g., 1996) has also suggested that cards that are *rejected* will be thought about very little, if at all (really this is just the corollary of the view that selected cards will be thought about in order to rationalise their selection). Evans (1996) also proposed that one way to provide evidence for the heuristic-analytic account would be to demonstrate that participants attend specifically to the cards that they end up selecting. Just such evidence was, in fact, uncovered in a study by Evans, Ball, and Brooks (1987), who used a computer presentation of the selection task with the full negations paradigm, where the order in which decisions were made about each card was recorded. Based on the assumptions of the heuristic-analytic theory, Evans et al. (1987) predicted that cards that were *considered* first would be those that were deemed relevant and subsequently selected. As predicted, a strong correlation was found between decision order and choice. However, it is possible that this result merely reflects a preference for people to make 'select' decisions before 'reject' decisions, and that what is really being shown in this decision-order paradigm is a response bias as opposed to an attentional bias (Roberts, 1998b).

A later study by Evans (1996) aimed to explore this effect further and claimed to be a stronger and more direct test of the heuristic-analytic theory. The two experiments that Evans reported both included a mixture of abstract and thematic versions of the selection task. Using a computer presentation of the tasks, participants were requested to point the mouse cursor at cards that they were considering choosing, but only to click the cursor on that card when they were sure they wanted to choose it. The computer recorded cumulatively the time spent pointing at each card. Evans refers to the latencies that were recorded as *card inspection times*. According to the heuristic-analytic theory participants should only attend to cards that they end up selecting. This is because—guided by linguistic cues (i.e., the '*if-heuristic*' and the '*matching-heuristic*')—they will only attend to *relevant* cards (i.e., those cards named in the rule). Heuristic, implicit processes lead to a decision to select what are perceived to be relevant cards, and participants then use explicit thought processes to rationalise those choices, which requires processing time. Evans (1996, p. 226), therefore, derived two main predictions about card selections and the time that people should spend looking at cards, as follows:

Prediction 1: Cards which are associated with higher selection rates will also be associated with longer inspection times.

Prediction 2: On any given card, those participants who choose the card will have higher inspection times than those who do not.

If, however, participants *were* analysing the consequences of turning each card, then, argued Evans (1996), the inspection-time paradigm should demonstrate equal looking times on all four cards. Evans' data for both of his reported experiments supported his predictions. Indeed, Evans found that differences in the times spent looking at selected cards compared to non-selected cards were sometimes as high as 13 seconds. Evans observed this inspection-time effect for both abstract and thematic versions of the selection task. He therefore claimed support for the heuristic-analytic theory; individuals focus their attention on heuristically-cued 'relevant' information, attend only to this information, and then go on to engage in non-consequential reasoning aimed at rationalising decisions based on perceived information relevance.

There are, however, grounds for caution before accepting Evans' (1996) conclusions. In particular, Roberts (1998b) has claimed that there could be a different explanation for Evans' results. In a series of five experiments with variants of the basic mouse-tracking and inspection-time paradigm, Roberts (1998b) first reduced, and then eventually eliminated—and even reversed—the inspection-time effect. Roberts claimed that a number of biases that derive from the mouse-tracking methodology itself could be responsible for the inspection-time effects demonstrated by Evans (1996). Roberts proposes that there are two key biases that are created by participants having to divide their attention between the selection task and the mouse-tracking task (which is used to show which card is under consideration). He calls these biases *sensory leakage* and *forgetting to move the mouse*. Sensory leakage means that cards could be viewed, considered and rejected before the mouse has had a chance to reach them. Participants could well be looking at all four cards for at least some of the time, but this information does not get 'logged' by the methodology. Forgetting to move the mouse could occur because participants may be so engaged in the problem that they simply forget to move the mouse to show what cards they are currently considering. These latter two task-format biases would result in inflated inspection times on cards that the mouse cursor was left on.

Roberts also suggests a third bias, *hesitation bias*, which is the tendency to hesitate before making an active decision. He argues that in the selection task participants are only required to make a decision about cards they are going to choose, and any pause before these choices could result in inflated times over the chosen cards. These three biases, then, would provide an artefactual explanation of the data presented by Evans (1996), as opposed to an explanation based on the theoretical claim that choices on this task are guided by heuristics processes and that the inspection-time effect arises from the analytic rationalisation of to-be-selected cards.

Roberts (1998b) was able to remove all three artefactual biases by making modifications to the task. After a replication of Evans' first experiment, Roberts attempted to remove sensory leakage and forgetting to move the mouse by ensuring that participants were only allowed to see one card at a time. Each card was 'blacked out', and pointing the mouse cursor over the required card allowed the participant to view the card. As this was done, the time was again recorded cumulatively. The evidence revealed a reduced inspection-time effect for this study. To eradicate the hesitation bias that is caused by only making active decisions, a third experiment required that participants make active choices over all four cards, either registering a 'yes' (select decision) or a 'no' (reject decision). This causes inspection times to be inflated over all four cards, thus counteracting the effect of the hesitation bias. Again, a reduced inspection-time effect was evident in this situation. The fourth experiment by Roberts used both individual card presentation and forced 'yes/no' decisions, so removing all three sources of bias. This resulted in the complete elimination of the inspection-time effect. Roberts' final experiment reversed the inspection-time effect using a *de-selection* task that required participants to make active decisions about cards that they *did not want to select*. The results showed the reverse of the inspection-time effect demonstrated in Experiment 1, where times were now longer for cards that were not selected than for those cards that were selected. Roberts (1998b) concludes that the inspection-time effect appears to be artefactual and is due to the sources of task-format bias that he identifies.

However, Evans (1998a) argues that there is still plenty of evidence to suggest that card selections are guided by heuristic processes (see Evans, 1984; Evans, Clibbens, & Rood, 1996) and that Roberts' experiments show that task manipulation, and not just relevance, can affect what is attended to on the selection task. Evans claims that Roberts' findings do not damage the heuristic-analytic theory, but instead call into question the use of the mouse-tracking methodology as a valid process-tracing technique. However, if the inspection-time paradigm could be utilised in a different way, for example, by using a more sensitive and direct measure of on-line attentional processing, then the paradigm might still prove to be a useful way to investigate the role of heuristic and analytic processes in reasoning. It is argued here that the inspection-time effect would lend itself extremely well to another well-known methodology for investigating the role of internal processes in cognition: that of eye-movement tracking.

2.3 Eye-movement Tracking and the Selection Task

Eye-movement tracking has been used extensively to investigate the on-line locus of attentional processing on many different tasks (e.g., Fisher, 1999; Klein, Kingstone, & Pontefract, 1992; Reuter-Lorenz & Frendrich, 1992). Recently, too, the methodology has been deployed very successfully in reasoning and problem solving contexts, for example, in the study of insight problem solving (Grant & Spivey, 2003; Knoblich, Ohlsson, & Raney, 2001), sentence-verification processes (e.g., Underwood, Jebbett, & Roberts, 2004), conditional inference (Schaeken, Fias, & d'Ydewalle, 1999; Schroyens, Schaeken, Fias, & d'Ydewalle, 2000), and syllogistic inference (Ball, Phillips, Wade, & Quayle, 2006; Espino, Santamaria, Meseguer, & Carreiras, 2005).

Although it is possible to move attention *covertly* (i.e., without moving the eyes), it is generally acknowledged that with visually-based stimuli it is far more efficient to move the eyes rather than merely to move attention (e.g., He & Kowler, 1992; Sclingensiepen, Campbell, Legge, & Walker, 1986). Moreover, there is substantial evidence indicating that attention actually *precedes* a saccade to a new location (Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995), that attentional movements and saccades are obligatorily linked (Deubel & Schneider, 1996), and that the length of a fixation or a gaze (which may involve two or more continuous fixations on the same location) provides a very good index of ease of processing (Liversedge, Paterson, & Pickering, 1998). In sum, in complex information-processing tasks such as reading or display-based reasoning, the coupling between eye location and the locus of ongoing attentional processing is likely to be very tight indeed (see Rayner, 1998, for further pertinent arguments).

Given that previous work suggests that eye-movement tracking may allow for a detailed investigation of on-line attentional processing, together with the fact that Evans' (1996) study clearly possessed a number of methodological problems, the present thesis set out to employ the eye-tracking methodology in order to examine on-line attentional processing in the selection task. A key motivation behind this research endeavour was to try to produce some novel process-tracing data that might help to inform the current theoretical understanding of reasoning processes on the abstract selection task.

2.4 Experiment 1

Experiment 1 aimed to employ the eye-movement tracking methodology within Evans (1996) inspection-time paradigm as a replacement for the more problematic mouse-tracking methodology. The use of eye-movement tracking as a measure of attentional processing should result in the eradication of all three sources of task-format bias identified by Roberts (1998b), thereby enabling the predictions of Evans' (e.g., 1996, 2006) heuristic-analytic theory to be investigated more directly. As participants did not need to use a mouse pointer to make their selections there was no longer the possibility of hesitation bias causing inflated times over active choices. Sensory leakage and forgetting to move the mouse were also no longer problematic, and, given the strong link between attentional processing and eye movements, there was little likelihood of participants looking at one card and thinking about another, or forgetting to move their eyes to a card that they were thinking about.

Experiment 1 used only abstract, indicative versions of the selection task within a full negations paradigm (cf. Evans et al., 1996). As such, participants were presented with a total of four selections task, with conditional rules being constructed by systematically permuting the presence of affirmative versus negative components in the antecedent and consequent cases of the rule. The four tasks were not presented by computer—as in the Evans (1996) and Roberts (1998b)—but on A4 cards in front of the participant, allowing eye movements to be calibrated and tracked throughout the trials. The eye tracker logged cumulative dwell times for each card. All four selection tasks were also shown to participants in a slightly non-standard format as presentation was constrained by the use of the eye tracker. The rules and instructions appeared as normal, but cards were presented in a 2 x 2 array, as was done in the study by Evans, Ball, and Brooks (1987).

The heuristic-analytic theory (Evans, 1996, 2006) predicts that participants should think more about the cards that they end up selecting than those that they end up rejecting because of the operation of analytic rationalisation processes in the case of the former but not the latter. Thus, in line with the heuristic-analytic theory it was expected that: cards that were more likely to be selected would have *longer inspection times* than cards that were less likely to be selected (Prediction 1); and that for any given card, those people who selected it would have longer inspection times than for those who did not select it (Prediction 2).

2.4.1 Method

Design

The design was principally a correlational one in which two types of measures were taken: (1) whether or not a card was selected; and (2) the associated inspection time for the card.

Participants

Participants were 34 undergraduates at the University of Derby who participated in the experiment in order to gain course credit. They had not had any teaching on the psychology of reasoning before this experiment.

Materials and apparatus

Reasoning tasks

The standard negations paradigm was used that permuted negative components through an indicative conditional rule of the basic form '*If p, then q*, to produce four rules (see Table 2.1 below). Participants were tested individually and all participants received all four problems. Each problem was presented at an angled table directly in the participant's line of vision. Participants sat at a chair that could be raised or lowered in order to get the best calibration. The angled table was located approximately 60 cm in distance from where the participant was sitting. The problems were presented individually on white A4 sheet with the conditional rule at the top of the page, a reminder of the instructions in the middle of the page, and the four cards in the lower part of the page in a 2 x 2 arrangement. The location of the cards in the four possible positions on the sheet was randomised for every presentation, as was the order of problem presentation.

TABLE 2.1

Four types of conditional rules used showing negated components and logical cases.

	Logical case			
Rule	***TA***	***FA***	***TC***	***FC***
If *A* then *3*	**A** *p*	J not-*p*	**3** *q*	7 not-*q*
If *E* then not *5*	**E** *p*	L not-*p*	2 not-*q*	**5** *q*
If not *S* then *9*	D not-*p*	**S** *p*	**9** *q*	4 not-*q*
If not *N* then not *8*	T not-*p*	**N** *p*	1 not-*q*	**8** *q*

TA = True antecedent, FA = False antecedent, TC = True consequent, FC = False consequent. Bold type indicates those cards that match those in the named rule. Logical choices for all four rules are TA and FC.

Eye tracking equipment

The eye-tracker that was used was the Applied Science Laboratories 4000 system. This system is video-based and uses a near-infrared light source and two video cameras. The 'scene' camera locates the participant in their environment, and the 'eye' camera produces a close-up image of one eye. The light source is guided through an arrangement of mirrors and lenses into the participant's eye and produces a 'retinal reflex' (which effectively makes the pupil appear bright) and a 'corneal image' reflection (a very bright smaller reflection off the front surface of the eye). The relationship between these two reflections changes as the participant's eye moves, and this relationship is used to calculate point-of-gaze coordinates. Using this system a participant's point-of-gaze can be determined to an accuracy of approximately one degree of visual angle.

Before point-of-gaze recording takes place, the relationship between the pupil and corneal reflections must be determined when the participant is looking at known points on the scene plane (the image on the scene camera). This is done using a 3 x 3 grid of nine evenly spaced points which are placed at the same location as the to-be-presented stimulus. The relationship between the corneal and pupil reflections is stored by the computer as the participant serially fixates on the nine calibration points. Using the data from the nine calibration points, it is then possible to interpolate the point-of-gaze across the whole scene.

During the experiment, the point-of-gaze was superimposed onto the scene video image as a small dark square. This was videotaped and later used in the data analysis along with a fixation data-file produced by the eye-tracking system that recorded time data and horizontal and vertical point-of-gaze coordinates, sampled at 50 times per second. In order to get accurate and detailed information about point-of-gaze, it was important to use this fixation data-file as well as the video scene record. An algorithm that considers both spatial and temporal characteristics was used to convert the raw point-of-gaze coordinates into fixations. To group the raw data into fixations the algorithm used a basic spatial threshold of one degree visual angle, with a minimum fixation duration of 100 msecs. This fixation algorithm was supplied with the ASL system and is equivalent to other algorithms that have been reported in the literature (e.g., see Nodine, Kundel, Toto, & Krupinski, 1992).

Procedure

Participants were tested individually. They were seated at a viewing distance of 60 cm from the presented materials and were asked to sit as still as possible whilst their eyes were calibrated using a 9-point presentation matrix. Once the participant's point-of-gaze coordinates had been determined, participants were presented with the general task instructions. These appeared separately from the four selection task themselves. Participants had as long as they wanted to read these general instructions, which were as follows:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. Each of these problems will appear on a separate sheet in front of you. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single figure number on the other side. Naturally only one side of each card will be visible to you.
>
> For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. When you make a decision in relation to each problem, please point briefly at the card or cards you feel need to be turned over using the hand-held pointer.

Once the participants had read through the instructions silently to themselves, the experimenter also read them aloud. This was done to ensure that the participant had time to grasp fully the requirements of the experiment. Each of the four selection-task problems was then presented one at a time in a random order. Participants indicated their selection responses with a 20 cm metal pointer, which was employed to avoid any interference to the eye-movement recording that may have occurred if participants had been required to raise their hand to the scene plane in order to use a finger to point at their card choices.

2.4.2 Results

Coding the inspection-time data

To enable use of the fixation data as a quantitative measure of eye-movement behaviour, the fixation data-file had to be considered alongside the scene-video record so that precise fixation data could be extracted from the point where the study material for each selection task was presented. To this end, a computer-based system was used, referred to as EMAT (Eye Movement Analysis Tool; Mugglestone, 1999), which allowed both the fixation data-file and the scene-video record to be linked together to ensure accurate temporal synchronisation. EMAT employs a Windows PC environment, and its basic components include: (1) a video recorder with a RS232 interface card (to allow video 'timecode' information to be accessed by the computer); (2) a personal computer with a video overlay card; and (3) bespoke software written in MS Visual Basic. EMAT allows the video scene to be viewed via a video playback window and the video recorder can be operated through the system.

When conducting the experiment, the calibration process and the calibration chart for each participant were recorded on to the video before the rest of the material was presented. The recording of the fixation data-file began at the same time as the video was set to record. This was marked on the video with a timecode overlaid onto the scene. Temporal synchronisation was achieved by locating the position on the videotape where the timecode started. This point was then used to create a link between the fixation data-file and the video scene. This link between the fixation data-file and the video scene was made anew for each participant.

The EMAT system allows for analysis of fixation data in relation to specific 'areas of interest' on the scene video, including, the *number of fixations* in any areas of interest and the *total fixation times* in those areas of interest. To achieve this a scaling procedure relates the coordinates of the fixation data-file to the pixel values in the scene-video playback window. This made it possible to superimpose the point-of-gaze information from the fixation data-file onto the video scene. The scene video already had a point-of-gaze cursor that was superimposed (by hardware) at the time of the experiment, and the fixation point plotted by the EMAT system was then expected to follow this original point-of-gaze cursor exactly. This was particularly useful as it was possible to see any errors in scaling or timing immediately, and problems could then be easily rectified.

To analyse the data from the current study, areas of interest were created by drawing around the four cards in the selection task. This created a template that was used throughout the coding stage. Whilst the fixation data-files and video scene were played together, the EMAT system produced a score of the total dwell time (sum of all fixations) on each of the four cards, for each problem, for each participant. These times were converted to an Excel file, and it was these total times that provided the 'card inspection time' data that were used in subsequent statistical analyses. Neither re-fixations nor first fixation points where measured, and this was the same for all eye-tracking experiments in the thesis. Note that for each selection task, the coding process was halted immediately after each participant had made their final decision, that is, at the instant that the pointer moved away from the last chosen card.

Statistical analysis

Card selection frequencies

The selection frequencies and the mean inspection times for all cards are shown in Table 2.2(b). The primary concern was to see whether or not the pattern of selections conformed to that found in the literature. Using negated components we should expect to find: (1) a clear matching effect on all four logical cases, and (2) a preference for TA over FA with TC and FC choices intermediate (although some studies report a verification bias as well as matching bias when negation is controlled, for example, see Krauth, 1982; Reich & Ruth, 1982).

Matching indices (as described by Evans, Ball, & Brooks, 1987) were computed for each participant across each rule by comparing the frequency of selections which match or mismatch [see Table 2.2(a)]. A one-tailed Wilcoxon test showed that more antecedent matching cards were selected than mismatching ones (p = .003), and that more matching consequent cards were selected than mismatching ones (p = .05). Mean percentage frequency of selections is shown in Table 2.2(b). A significant difference was found on a Friedman test for the frequency of choices of the four logical cases, (χ = 34.02, df = 3, $p < .001$). As the literature suggests, there was a large preference for TA choices over all other cases, and FA was the least popular choice. There was also a preference, although smaller, for TC over FC, which as suggested above could be evidence for a verification bias as well as matching bias.

The selection data are therefore typical of what are usually observed in the negations paradigm.

Card inspection times

An exploratory investigation of the inspection-time data showed that they would not be suitable for parametric statistical analysis as they were positively skewed. In this situation a logarithmic transformation is an appropriate transformation to perform (Miles & Shevlin, 2001). A constant of 0.2 was added to each score prior to its transformation. (This problem was encountered for Experiments 1 to 4 and so the same transformation was used throughout). For clarity of interpretation, means after transformation but converted back to their original units are reported where applicable.

Prediction 1 states that cards with higher selection frequencies will have higher mean inspection times associated to them. There should, therefore, be a correlation between the frequency with which a card is selected and the mean inspection time for that card. Mean inspection times and the percentage frequency of card selections are shown in Table 2.2(b). A correlation analysis showed that there was indeed a strong association between selection frequencies and inspection times: $r = .55$, $N = 16$, $p = .028$ (transformed data).

The second analysis tested Prediction 2: that for each card on each rule the participants who chose that card would spend more time looking at it than participants who did not. For this analysis, Evans (1996) computed point biserial correlations between selections (1 to indicate selected cards or 0 to indicate non-selected cards) and the inspection times across all participants, and found that for each of the 16 correlations there were significantly more positive correlations than would be expected by chance.

TABLE 2.2

(a) Formulae for comparing matching indices across (i) antecedent cases and (ii) consequent cases; (b) Percentage frequency of card selections and overall mean inspection times (seconds) and log times (seconds) for each item in Experiment 1 (*N=34*).

(a)

Matching indices - one-tailed Wilcoxon comparisons across:

(i) matching antecedent (TA:1 + TA: 2 + FA: 3 + FA:4) versus mismatching antecedent (TA:3 + TA:4 + FA:1 + FA:2)

(ii) matching consequent (TC:1 + TC:3 + FC:2 + FC:4) versus mismatching consequent (TC2: + TC4: + FC1: + FC:3).

(b)

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	88.2		8.8		35.3		14.7	
	Natural data: time	1.42	1.22	0.42	0.59	1.29	1.27	0.75	0.96
	Transformed data: log time	0.10	0.32	-0.34	0.32	0.03	0.36	-0.19	0.38
	Transformed data: corrected time	1.06		0.26		0.87		0.45	
2. If p then not q	Frequency of selections	70.6		20.6		20.6		29.4	
	Natural data: time	2.19	2.18	1.14	2.32	0.98	1.73	0.59	0.20
	Transformed data: log time	0.21	0.41	-0.12	0.42	-0.16	0.41	0.10	0.37
	Transformed data: corrected time	1.42		0.56		0.49		1.06	
3. If not p then q	Frequency of selections	52.9		23.5		64.7		29.4	
	Natural data: time	1.39	1.52	1.50	1.06	2.26	2.39	0.98	1.05
	Transformed data: log time	0.02	0.41	0.16	0.25	0.22	0.40	-0.04	0.30
	Transformed data: corrected time	0.85		1.25		1.46		0.71	
4. If not p then not q	Frequency of selections	55.9		35.3		50.0		32.4	
	Natural data: time	1.69	2.12	1.91	2.00	1.58	2.20	1.92	1.69
	Transformed data: log time	0.07	0.43	0.10	0.48	0.02	0.44	0.17	0.39
	Transformed data: corrected time	0.98		1.06		0.85		1.28	
Mean		66.9		20.1		42.7		26.5	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

It is recognised here, however, that the use of the correlation is unsatisfactory—as also noted by Roberts (1998b)—the problem being due to the imbalance in numbers of selected and non-selected cards. Roberts suggests a slightly different analysis to see whether for each of the 16 cards the mean inspection time is greater for the people who select a particular card than for those who do not. Table 2.3(a) shows the mean inspection times for selected and non-selected cards for all 16 cases. On the transformed data, the difference between mean times on selected and non-selected cards was in the expected direction for 15 of 16 cards (correlations range from -0.08 to 0.60), $p = .001$, two-tailed with the binomial test.

Roberts (1998b) argues that one potential problem with the Prediction 1 and Prediction 2 analyses is that they lack statistical power because they are *item analyses*. He therefore attempts a more powerful analysis of the data at the participant level, and makes the following prediction:

> Prediction 3: For each participant, mean inspection times should be longer for the cards that they have selected than for those that they have not selected.

For each individual, two means were calculated from the transformed data and a within-subjects analysis of variance (ANOVA) was performed [see Table 2.3(b)]. This confirmed Prediction 3, that inspection times were longer for the cards that were selected than those that were not, $F(1, 33) = 65.13$, $MSE = 2.08$, $p < .001$.

TABLE 2.3

Mean inspection times for selected and non-selected cards, for Experiment 1 showing (a) inspection times (in seconds) for each item, and b) inspection times (in seconds) and log times (in seconds) overall by participants ($N = 34$).

(a) by items

		Selected			*Not Selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	30	1.28	1.55	4	0.24	0.43
	FA	3	1.00	1.37	31	0.22	0.33
	TC	12	1.18	1.64	22	0.76	1.11
	FC	5	1.84	2.10	29	0.33	0.52
2. If p then not q	TA	24	2.09	2.80	10	0.49	0.73
	FA	7	1.58	1.81	27	0.42	0.97
	TC	7	1.54	2.41	27	0.35	0.61
	FC	10	1.58	0.51	24	0.90	0.63
3. If not p then q	TA	18	1.58	2.13	16	0.38	0.56
	FA	8	1.12	1.30	26	1.28	1.56
	TC	22	1.94	2.75	12	0.87	1.36
	FC	10	1.22	1.66	24	0.56	0.70
4. If not p then not q	TA	19	1.54	2.30	15	0.52	0.93
	FA	12	1.75	2.67	22	0.80	1.50
	TC	17	1.46	2.27	17	0.46	0.88
	FC	11	2.14	2.70	23	0.98	1.54

(b) by participants

	Selected		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data	2.16	1.40	0.90	0.65
Transformed data: log time	0.24	0.24	-0.11	0.27
Transformed data: corrected time	1.54		0.58	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent, FA = false antecedent, TC = true consequent, FC = false consequent.

Despite finding support for the three predictions, one possible criticism for interpreting these findings as support for Evans' heuristic-analytic theory, is that although there is evidence that people select more matching than mismatching cards, together with evidence that people spend longer inspecting selected cards than rejected cards, the analysis doesn't show that these two findings hold conjointly (i.e., that people who select matching cards are the same people as those that spend longer looking at selected cards). The heuristic-analytic theory would predict reliable differences between selected and non-selected inspection times for both matching and mismatching cards. Therefore, it is possible to derive two more predictions from the theory:

Prediction 4(a) For each participant the mean inspection time will be longer for the matching cards they select than for those they reject; and

Prediction 4(b) For each participant, the mean inspection time will be longer for the mismatching cards they select than for those they reject.

Planned contrasts were carried out and provided good support for both predictions. For the matching-selected versus matching-rejected comparison, the respective mean inspection times were 1.71 s and 0.82 s (transformed data in natural units) which was highly reliable, $F(1, 24) = 47.98$, $MSE = 1.84$, $p < .001$. For the mismatching-selected versus the mismatching-rejected comparison, the mean inspection times were 1.50 s and 0.43 s (transformed data in natural units), which was also highly reliable, $F(1, 24) = 59.14$, $MSE = 4.59$, $p < .001$.

2.4.3 Discussion

The results of Experiment 1 appear to provide good support for the three predictions derived from Evans (1996, 2006) heuristic-analytic theory. The results demonstrate that cards that are selected are inspected for longer than cards that are not selected. Utilising the eye-tracking methodology has eliminated the sources of bias identified by Roberts and has left a robust inspection-time effect. It is noted, however, that the present inspection times are relatively shorter than the times reported by Evans (1996). This is important because Evans makes claims about the processes involved during the time participants spend looking at cards. It is possible that extreme outliers in Evans' (1996) study could account for the longer-duration inspection-time data that he obtained (see Roberts, 1998b, for related arguments).

It also seems that Prediction 4 continues to demonstrate good support for the heuristic-analytic theory. It is useful to note that Evans (1996, 1998a) has clarified that 'relevance' effects can extend well beyond matching cards in determining selections. For example, on the abstract task Evans (1989) proposed the *if-heuristic* to drive the selection of the true antecedent card across all rules regardless of whether or not these cards have matching status. He has also found that when selection patterns vary—as they do across thematic versions of the selection task—they can still be interpreted as arising from relevance judgements (Evans, 1996). So, when people make different card selections from one another, all that is happening is that different linguistic, pragmatic or attentional factors have cued relevance via heuristic processes.

Before accepting the findings of Experiment 1 as definitive evidence for the heuristic-analytic theory, however, it is important to be aware of a further possible bias arising from the design of this experiment. In this experiment the task involved participants pointing to the cards that they wanted to choose. The total inspection time on a selected card, therefore, included both: (1) the time spent looking at the card during its consideration by the reasoner, *and* (2) the pointing-response time—as the task only required participants to make active responses (i.e., pointing) for cards that they wanted to select. This procedural requirement, therefore, could have had the same effect on inspection times as the *hesitation bias* that Roberts (1998b) described and dealt with in his Experiment 3 (i.e., it could have inflated the inspection times for selected cards over rejected ones). The concern, then, is that the reliable inspection-time findings observed in Experiment 1 could, once again, be a consequence of a methodological artefact that has inadvertently confounded the results (I am grateful to Max Roberts for comments and discussions that led to the identification of this problem).

An artefact-based explanation of the findings of Experiment 1 could be rejected if it could be demonstrated that any time inflation arising from an *active* decision requirement would be inadequate, in itself, to account for the full difference in inspection times between selected and non-selected cards. On examination of the video files for instances when card selections occurred, it could be noted that the actual time taken from when a participant initiated the movement of the pointing device to the time when they registered the card selection was always very short (usually around 0.5 s and seldom taking longer than 1 s). In contrast, the mean inspection-time difference between selected cards and non-selected cards was 1.26 s (natural data); 0.96 s for transformed data in original units), which indicates that around 0.5 s of the inspection-time difference between selected and non-selected cards could *not* be accounted for by a pointing artefact. Therefore, this small, residual inspection-time effect may well be a reflection of the predicted influence of an analytic rationalisation process.

2.5 Experiment 2

Although Experiment 1 seems to provide support for an inspection-time effect as predicted by Evans (e.g., 1996, 2006), the support is weakened due to the identified pointing bias that may arise from the fact that participants were only required to make active choices for cards that they wanted to select. The purpose of Experiment 2 was to remove the pointing bias by fully separating the process of registering the selection of cards from the process of reasoning on the task.

For each presented selection task in Experiment 2, then, participants were instructed to press a button to activate a light (situated above the problem) that indicated when they were ready to make their decisions. This light was visible on the recording of the participant's eye movements, and so at the data-analysis stage it was possible to omit the pointing time from the cumulative inspection-time scores for selected cards.

If the inspection-time effect on selected cards observed in Experiment 1 was purely due to this pointing bias then a reduced (or possibly completely eliminated) inspection-time effect would be expected in Experiment 2. However, if the inspection-time effect remained despite eradicating the pointing bias, then this would lend very good support to the heuristic-analytic account of reasoning on the selection task as proposed by Evans (e.g., 1996, 2006).

2.5.1 Method

Participants

Participants were 30 undergraduates at the University of Derby who participated in the experiment in order to gain course credit. They had not had any teaching on the psychology of reasoning before this experiment.

Materials and apparatus

All materials and apparatus were identical to those used in Experiment 1. Additionally, participants were given a small button-box to hold in their non-favoured hand. Before making their selection decisions participants were required to press the button, which briefly activated a light to indicate that they were ready to make their choices.

Procedure

The procedure was identical to that used in Experiment 1, except that the instructions were modified slightly to incorporate the use of the button-box:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. Each of these problems will appear on a separate sheet in front of you. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single figure number on the other side. Naturally only one side of each card will be visible to you.
>
> For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. Once you have reached a point where you think you know which card or cards need to

be turned over then please press the button on your hand-held button-box. This will momentarily activate a light so that the experimenter knows that you are ready to indicate your selections. The experimenter will then ask you to point at the card or cards that you feel need to be turned over using the hand-held pointer.

As in Experiment 1, the experimenter went through the instructions with the participants to ensure that they understood them. Again each task was presented separately. Responses were made using a 20 cm metal pointer.

2.5.2 Results

Coding the inspection-time data

The process of data coding was identical to that of Experiment 1. However the pressing of the button activated a small light above each of the selection-task problems that were presented. At the coding stage, when this light appeared the coding of the data was halted. This meant that unlike Experiment 1, inspection times now did not include the time it took participants to make their decisions. The pointing bias was, therefore, fully eliminated in the analysis of data for Experiment 2.

Statistical analysis

Card selection frequencies

Matching indices again showed significant effects in line with the matching bias hypothesis. A one-tailed Wilcoxon test revealed more antecedent matching cards were selected than antecedent mismatching ones (p = .02) and more consequent matching cards were selected than mismatching ones (p = .004). Again, mean percentage of frequency can be seen at the bottom of Table 2.4. A significant difference was found for the frequency of choices of the four logical cases (χ = 44.17, df = 3, $p < .001$, Friedman test). Again, there was a preference for TA choices over all other cases with FA being the least popular choice. There was also a preference for TC over FC, which again could indicate a verification bias.

Card inspection times

Analyses of the inspection-time data were identical to those of Experiment 1 using log-transformed data. The frequency of selections and mean inspection times for each card can be seen in Table 2.4. The correlation between selection frequency and mean inspection time was significant, $r = .62$, $N = 16$, $p = .007$.

TABLE 2.4

Percentage frequency of card selections and overall mean inspection times (in seconds) and log times (in seconds) for each item in Experiment 2 ($N = 30$).

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	86.7		6.7		60.0		3.3	
	Natural data: time	3.73	3.80	1.67	1.67	3.57	3.19	2.21	2.64
	Transformed data: log time	0.43	0.40	0.07	0.47	0.44	0.37	0.13	0.49
	Transformed data: corrected time	2.49		0.98		2.55		1.15	
2. If p then not q	Frequency of selections	83.3		23.3		13.3		43.3	
	Natural data: time	5.04	4.75	3.04	5.26	2.39	4.02	4.74	6.00
	Transformed data: log time	0.55	0.41	0.24	0.44	0.18	0.43	0.42	0.50
	Transformed data: corrected time	3.35		1.54		1.31		2.43	
3. If not p then q	Frequency of selections	53.3		16.7		56.7		33.3	
	Natural data: time	5.98	11.15	5.44	7.59	5.70	7.67	4.30	5.87
	Transformed data: log time	0.43	0.53	0.54	0.42	0.51	0.50	0.43	0.47
	Transformed data: corrected time	2.49		3.27		3.04		2.49	
4. If not p then not q	Frequency of selections	60.0		30.0		50.0		36.7	
	Natural data: time	4.46	3.57	5.12	4.91	5.15	5.82	7.15	9.73
	Transformed data: log time	0.53	0.38	0.55	0.42	0.53	0.41	0.60	0.49
	Transformed data: corrected time	3.19		3.35		3.19		3.78	
Mean		57.3		19.2		45.0		29.2	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent

Table 2.5(a) shows the mean inspection times (before and after transformation) for selected cards and non-selected cards for each of the 16 cards. For transformed data the difference between mean inspection times on selected and non-selected card times was in the expected direction for 13 of the 16 cases (correlations range from -0.09 to 0.49) which is significant on the binomial test, $p = .004$, two-tailed.

Despite having removed the biases identified by Roberts and in Experiment 1, there was still a reliable inspection-time effect. This was confirmed using a participant-level analysis where the pairs of mean inspection times [see Table 2.5(b)] for each participant for selected and non-selected cards were compared using ANOVA, $F(1, 29) = 43.25$, $MSE = 0.98$, $p < .001$.

Again, two separate contrasts (by-participants) were undertaken to determine the reliability of the selected versus non-selected difference for matching cards (3.25 s vs. 2.43 s; transformed data in natural units), and for mismatching cards (3.60 s vs. 1.31 s). Both contrasts were reliable, $F(1, 23) = 5.06$, $MSE = 0.41$, $p = .034$, and $F(1, 23) = 41.38$, $MSE = 3.92$, $p < .001$, respectively.

TABLE 2.5

Mean inspection times for selected and non-selected cards, for Experiment 2, showing (a) inspection times (in seconds) for each item, and b) inspection times (in seconds) and log times (in seconds) overall by participants ($N = 30$).

a) by items

		Selected			*Not Selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	26	2.43	3.29	4	2.82	6.56
	FA	2	4.93	4.98	28	0.85	1.43
	TC	18	2.62	3.23	12	2.49	4.09
	FC	1	10.98	10.98	29	1.06	1.91
2. If p then not q	TA	25	3.69	5.41	5	2.14	3.18
	FA	7	3.43	5.35	23	1.18	2.34
	TC	4	1.42	1.53	26	1.28	2.52
	FC	13	3.35	6.40	17	1.89	3.47
3. If not p then q	TA	16	3.97	9.72	14	1.50	1.69
	FA	5	3.04	4.58	25	3.35	5.61
	TC	17	3.69	4.73	13	2.37	6.97
	FC	10	4.37	7.39	20	1.84	2.76
4. If not p then not q	TA	18	4.59	5.67	12	1.84	2.66
	FA	9	2.96	3.82	21	3.60	5.67
	TC	15	4.07	5.94	15	2.55	4.36
	FC	11	4.17	8.37	19	3.52	6.44

b) by participants	*Selected*		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data	5.33	4.61	3.34	3.39
Transformed data: log time	0.56	0.29	0.31	0.33
Transformed data: corrected time	3.43		1.84	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent, FA = false antecedent, TC = true consequent; FC = false consequent

2.5.3 Discussion

The results of Experiment 2 show that despite removing all sources of task-format biases—including the pointing bias identified in relation to Experiment 1—there remained a strong inspection-time effect, as predicted by Evans' (1996, 2006) heuristic-analytic theory. One interesting point of difference between the inspection-time effect seen in Experiment 2 and Experiment 1, however, is that the mean, participant-level, inspection-time difference between selected and non-selected cards in Experiment 2 (1.99 s for natural data; 1.59 s for transformed data in original units) was actually *greater* than that in Experiment 1, where the pointing bias was potentially present (1.26 s for natural data; 0.96 s for transformed data in original units). One cause of this finding might be the response-collection method used in Experiment 2, whereby participants alerted the experimenter before making their decisions. It is possible that this technique actually served to *amplify* the size of the inspection-time effect, so that participants repeated the processing of the cards so as to be sure of their selections before committing themselves. This explanation does not, of course, undermine the evidence for the predicted inspection-time imbalance that occurs between selected and non-selected cards, as the amplification of the effect can only occur if the effect is there in the first place.

The presence of the inspection-time effect in Experiment 2 is very different to that of Roberts (1998b) in his Experiment 3 (forced 'yes/no' choice on each card) and his Experiment 4 (forced 'yes/no' choice on each card combined with individual card presentation). Roberts' task modifications addressed the problem of hesitation bias and caused the inspection-time effect firstly to be reduced and then to be lost completely. By requiring participants to attend to each card individually, Roberts argued that the perceived relevance of the cards may have been altered, causing each one to appear equally relevant, so reducing the inspection-time effect. Roberts suggests that Evans' (1996) inspection-time effect is weakened because there is no competition between other information as there is when the cards are presented together. Once this method is combined with a forced 'yes/no' choice on each card, then participants are further encouraged to think about the alternative values of each card.

It is possible, then, that the inspection-time effect as predicted by the heuristic-analytic theory is, indeed, non-existent, or, alternatively, that the way in which Roberts has altered the selection task has disrupted the normal processes on the task and caused his null results. However, if an inspection-time effect was found with forced decisions for each card using the eye-movement methodology, this would go against the view espoused by Roberts, and would instead, be good support for the heuristic-analytic theory. It could also mean that Roberts' findings were not due to the non-standard task-format but rather because the mouse-pointing methodology lacks the sensitivity to detect inspection-time effects. In other words, mouse pointing may simply be a very poor process-tracing method that is unable to detect inspection-time effects, unlike eye-movement tracking, which affords a greater sensitivity in measuring momentary fluctuations in the allocation of attention to cards. In order to assess these ideas, forced select/reject decisions were required for all cards within the selection-task study that was run in Experiment 3.

2.6 Experiment 3

The aim of Experiment 3 was to investigate the use of the forced-decision task, where select/reject choices were enforced for all presented cards. Experiment 3 was, therefore, a replication of Roberts' (1998b) Experiments 3 and 4, which were designed to standardise any influence of the hesitation bias that is caused by participants making active decisions only about those cards that they want to choose, as having to make an active decision for all cards avoids any localised inflation of inspection-times for just selected cards that might arise with the standard presentation of the selection task. Roberts required participants to make an active decision about each card by adding a 'yes' and a 'no' button under each card. He found that the inspection-time effect was considerably reduced in this paradigm, only finding support for the more powerful by-participants analysis (Prediction 3) in that mean inspection times for selected cards were higher than mean inspection times for non-selected cards.

Evans (1998a) agrees that on this version of the task participants are bound to consider each card, and therefore it would be expected that the inspection-time effect would be weakened. He argues, however, that being forced to attend to a card that would normally be ignored does not mean that cards will necessarily be perceived as more relevant. Evans argues that although attention and relevance are compatible concepts they are not the same thing, citing evidence of the matching bias effect on truth-table tasks to support his claim. Evans (1972) showed that on a construction version of the truth-table task, participants show a preference to construct cases that match the lexical content of the rules. However, later studies employed the evaluation version of the task where participants are presented with cases that conform to the rule, contradict it or are irrelevant to the rule. Evans (1975) notes that this change of task makes no difference to participants' choices; they still exhibit matching bias although their attention has been drawn to cases that they would not normally consider.

Using the 'yes-no' decision technique is also another way to negate the influence of a possible pointing bias as identified in Experiment 1. Within the 'yes/no' paradigm, any inflation in the allocation of attention to a card that results from the need to point at it in order to register a decision would be equal for all cards. In terms of predictions, then, it was hypothesised—in line with the heuristic-analytic theory—that an inspection-time effect would still emerge in Experiment 3, even with the use of the 'yes-no' paradigm. Although this could be seen as a risky prediction, it was believed that the sensitivity of the eye-tracking method should enable the detection of small but reliable differences in inspection-times across cards. If this occurred, then this finding would seem to provide very good support for the heuristic-analytic account.

2.6.1 Method

Participants

Participants were 31 undergraduates at the University of Derby who participated in the experiment in order to gain course credit. They had not had any teaching on the psychology of reasoning before this experiment.

Materials and apparatus

All materials and apparatus were identical to those used in Experiments 1, but below every card two separate boxes labelled 'yes' and 'no' (separated horizontally by 0.8 cm) were placed at a distance of about 1 cm from the card's lower edge.

Procedure

The procedure was identical to that used in Experiments 1, except that participants were required to make an active decision about every card presented to them. The instructions reflected the changes to the task:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. Each of these problems will appear on a separate sheet in front of you. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single figure number on the other side. Naturally only one side of each card will be visible to you.
>
> For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. If you decide that a card needs to be turned over then please point to the 'yes' box under the card. If you decide that a card doesn't need to be turned over then point to the 'no' box below the card. You will need to make a decision about each of the cards presented to you.

As in Experiments 1 and 2, the experimenter went through the instructions with the participants to ensure that they understood them. Again, each task was presented separately and responses were made using a 20 cm metal pointer.

2.6.2 Results

Coding the inspection-time data

The same procedure was used for coding the data as in Experiments 1 and 2. Coding was halted when participants had finished making decisions about all four cards, that is, when the pointer was moved away from the final card on which a decision had been made.

Statistical analysis

Card selection frequencies

Support was again found for the matching-bias hypothesis as the matching indices showed significant effects. A one-tailed Wilcoxon test showed that more antecedent matching cards were selected than mismatching ones ($p = .004$) and more consequent matching cards were selected than mismatching cards ($p < .001$). Mean percentage frequency data are shown at the bottom of Table 2.6. A significant difference was again found for the frequency of choices of the four logical cases ($\chi = 31.45$, df = 3, $p < .001$, on a Friedman test). As with the previous two experiments there was a clear preference for TA choices over all other cases. FA was again the least popular choice and there was again a preference for TC over FC choices indicating a possible verification bias.

Card inspection times

Log-transformed data were analysed as in Experiments 1 and 2. The frequency of selections and mean inspection times for each card can be seen in Table 2.6. Identical analyses of the inspection time data as in Experiment 1 and 2 were performed. The correlation between selection frequency and mean inspection time was significant, $r = .61$, $N = 16$, $p = .011$ (transformed data).

TABLE 2.6

Percentage frequency of card selections and overall mean inspection times (in seconds) and log times (in seconds) for each item in Experiment 3 ($N = 31$).

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	90.3		16.1		74.2		16.1	
	Natural data: time	2.19	1.53	1.67	1.68	1.68	1.44	1.30	0.92
	Transformed data: log time	0.31	0.25	0.15	0.33	0.18	0.27	0.09	0.30
	Transformed data: corrected time	1.84		1.21		1.31		1.03	
2. If p then not q	Frequency of selections	90.3		29.0		16.1		64.5	
	Natural data: time	3.46	3.78	2.31	2.51	2.84	3.42	3.07	2.46
	Transformed data: log time	0.39	0.38	0.25	0.38	0.28	0.43	0.41	0.31
	Transformed data: corrected time	2.26		1.59		1.71		2.31	
3. If not p then q	Frequency of selections	61.3		25.8		83.9		22.6	
	Natural data: time	2.65	2.50	2.45	2.11	2.90	2.56	2.33	2.16
	Transformed data: log time	0.32	0.36	0.28	0.39	0.36	0.34	0.25	0.38
	Transformed data: corrected time	1.89		1.71		2.09		1.58	
4. If not p then not q	Frequency of selections	48.4		51.6		45.2		51.6	
	Natural data: time	1.96	1.57	2.61	2.34	2.08	1.55	2.56	2.02
	Transformed data: log time	0.22	0.34	0.30	0.38	0.24	0.35	0.34	0.31
	Transformed data: corrected time	1.50		1.80		1.54		1.99	
Mean		72.6		30.6		54.9		38.7	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

Mean inspection times (before and after transformation) for selected cards and non-selected cards for each of the 16 cards are given in Table 2.7(a). For transformed data the difference between mean inspection times for selected and non-selected cards was in the expected direction for 9 out of the 16 cases (correlations range from -0.14 to 0.74) which is non-significant on the binomial test, $p =$ .80 (two-tailed). However, the more powerful individual-level analysis across the pairs of mean inspection times [see Table 2.7(b)] for each participant for selected and non-selected cards was undertaken using ANOVA, and revealed a significant difference across cards, $F(1, 30) = 14.87$, $MSE = .11, p = .001$.

TABLE 2.7

Mean inspection times for selected and non-selected cards, for Experiment 3, showing (a) inspection times (in seconds) for each item, and b) inspection times (in seconds) and log times (in seconds) overall by participants ($N = 31$).

a) by items

		Selected			*Not Selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	28	1.80	2.20	3	2.09	2.09
	FA	5	1.99	2.48	26	1.12	1.51
	TC	23	1.28	1.57	8	1.54	2.01
	FC	5	0.78	0.90	26	1.09	1.38
2. If p then not q	TA	28	2.31	3.56	3	1.84	2.51
	FA	9	3.11	4.26	22	1.18	1.52
	TC	5	2.89	3.16	26	1.54	2.78
	FC	20	2.20	2.55	11	2.68	4.02
3. If not p then q	TA	19	2.04	3.10	12	1.66	1.93
	FA	8	1.42	1.91	23	1.80	2.63
	TC	26	2.31	3.00	5	1.31	2.38
	FC	7	1.31	1.90	24	1.71	2.45
4. If not p then not q	TA	15	1.54	1.75	16	1.39	2.18
	FA	16	2.20	3.26	15	1.46	1.91
	TC	14	1.35	1.84	17	1.71	2.27
	FC	16	2.20	2.84	15	1.75	2.26

b) by participants	*Selected*		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data	2.74	1.63	2.10	1.27
Transformed data: log time	0.33	0.22	0.25	0.21
Transformed data: corrected time	1.94		1.58	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

Two separate planned contrasts (by-participants) were performed in order to assess the reliability of the selected versus non-selected difference for matching cards (2.20 s vs. 1.71 s; transformed data in natural units), and for mismatching cards (2.09 s vs. 1.46 s). Both contrasts were reliable: $F(1,19) = 5.82$, $MSE = 0.28$, $p = .026$, and $F(1,19) = 11.53$, $MSE = 0.53$, $p = .04$, respectively.

2.6.3 Discussion

After removing all three task-format biases identified by Roberts (1998b), that is, sensory leakage, forgetting to move the mouse, and hesitation prior to registering active decisions, Experiment 3 still provided support for Evans' (1996) Prediction 1 (i.e., that cards with higher mean selection frequencies will also have higher mean inspection times). Thus, there was still a strong association between selection rates and time spent inspecting the cards. Evans' (1996) Prediction 2 (i.e., that for each card people who choose it will have longer inspection times than those who do not choose it) was not supported, but the more powerful individual-level analysis (Prediction 3) proposed by Roberts (1998b) was supported: overall, the cards that individuals select have longer inspection times than those cards that they do not select.

Despite creating a situation where participants are forced to look at all four cards and whereby any influences of active response requirements were equalised across all cards, it appears that the focus of participants' attention has actually only been affected very slightly relative to what was observed in Experiment 1 and 2, that is, there is just a small decrease in the size of the inspection-time effect overall compared to the previous two experiments. In particular, the participant-level data for Experiment 3 reveal the smallest inspection-time imbalance between selected and rejected cards across the three experiments (0.64 s for natural data; 0.36 s for transformed data in original units). However, this occurrence is not very surprising when Evans' (1998a) argument is considered: that just because participants are forced to look elsewhere on the task does not necessarily mean that heuristic processes are not involved in directing attention (see Evans, 1998a, and Roberts, 1998a, for related arguments). The key point is that even this *response-compelled attention* (as Roberts, 1998a, describes it) does not seem to undermine the presence of the inspection-time effect. Taken together, the results of Experiment 3 provide good support for the heuristic-analytic account of selection-task performance.

2.7 General Discussion of Experiments 1, 2 and 3

Evans' (e.g., 1996, 2006) heuristic-analytic theory of the people's selections on the abstract selection task claims that preconscious, heuristic processes direct attention towards cards which appear relevant (which end up being selected) and away from cards that seem irrelevant (which in turn are rejected), whilst conscious analytic processes only serve to *rationalise* decisions that have already been made on the basis of relevance. Experiments 1, 2 and 3 aimed to test the predictions of the heuristic-analytic theory, which proposes that an inspection-time effect should be observable when people engage in the task, that is, people should consider cards that they end up selecting for longer than those that they end up rejecting because of the operation of analytic rationalisation processes that are almost solely aimed at justifying heuristically-determined choices on this task.

The three experiments used a direct measure of on-line attentional processing, that of eye-movement tracking, which is a more sensitive, moment-by-moment index of the locus of participants' attentional focus during the task. As well as being a more sensitive method than previously used by Evans (1996) and Roberts (1998b), the eye-tracking methodology also allowed for the removal of task-induced biases that Roberts (1998b) has suggested may be the sole reason for the inspection-time effect that Evans (1996) observed.

In Experiment 1, a standard selection-task paradigm was employed. In this version active, pointing responses were registered for selected cards only. Although a robust inspection-time effect was observed, the required pointing response for only selected cards meant that the experimental condition may have inadvertently introduced a possible methodological artefact, in that the time taken to point to selected cards could have caused inflated inspection-times on these cards compared to non-selected ones. Experiments 2 and 3 were designed to eradicate such a pointing bias. Experiment 2 separated the reasoning part of the task from the selection of cards by monitoring inspection-times only up to the moment directly prior to participants making their decisions. In Experiment 3, the pointing bias was *equalised* across all four cards since participants were asked to make a yes or no decision for every card on the task. Both Experiments 2 and 3 established reliable inspection-time effects, although the effect in Experiment 3 was slightly reduced in magnitude.

The inspection-time effect, as predicted by Evans' (1996, 2006) heuristic-analytic account of the selection task, appeared to be highly reliable across Experiments 1 to 3, despite differences in the problem formats and decision requirements. A failure to support the inspection-time effect with the eye-tracking methodology could have caused serious problems for the heuristic-analytic account, and could have bought into doubt the theoretical assumptions of the theory (Roberts, 1998a, 1998b). The fact that the present dataset support the validity of an inspection-time effect seem to attest to the validity of Evans' heuristic-analytic predictions.

It is also noteworthy that despite the methodological problems that might have undermined Evans' (1996) mouse-tracking and inspection-time findings, a subsequent study by Roberts and Newton (2001) reported three new selection task experiments that added methodological improvements to the basic mouse-tracking approach. These new experiments indicated that mouse-pointing measures *can* be sensitive to effects predicted by the heuristic-analytic framework. In particular, one modification involved the use of 'change' tasks, whereby participants were given cards that were either presented as selected or non-selected and they had to change them where necessary. Results demonstrated a reliable association between card selection and increased inspection times. Roberts and Newton (2001) accept that this result does provide support for the view that heuristic-induced biases influence choices.

Roberts and Newton (2001) also presented two further studies (Experiments 2 and 3) using a rapid-response selection task (requiring a card decision within 2 s of its presentation) that led them to propose an important caveat concerning the adequacy of the heuristic-analytic theory—although they remained broadly favourable toward this account. They note that their rapid-response tasks *raised* levels of matching for consequent cards on certain rule forms (without changing levels of logical responding) in comparison with free-time tasks. This suggests that analytic processing arising in free-time situations may serve to overturn candidate cards cued through attentional heuristics, in contradiction to Evans' heuristic-analytic account, but in line with mental models proposals (see also Feeney & Handley, 2000; Handley, Feeney, & Harper, 2002, for other evidence claiming the involvement of models-based analytic processing in variants of the abstract selection task). It is possible, however, that the analytic effects that influence card selections identified by Roberts and Newton (2001) may be restricted to a subset of individuals, with a majority responding equivalently under both speeded and unspeeded conditions. Thus the heuristic-analytic theory may capture the behaviour of most individuals, whilst other accounts (e.g., mental models theory) may better describe the processing of a subset of individuals (see Stanovich & West, 1998, for evidence of individual differences in responding on the selection task).

One important question is whether *other* contemporary theories of the selection task can incorporate the reliable inspection-time effect found in these experiments. The first account to consider is Oaksford and Chater's (1994) information gain theory. It seems that this theory would have difficulty in dealing with the inspection-time effect as the account suggests that people derive expected information gains for each card, and this should presumably take an equivalent amount of time for each card whether a card ends up being selected or rejected. However, it is somewhat improper to critique this theory in relation to its apparent inability to explain the findings of inspection-time experiments because as it currently stands, the information gain account is a *computational-level theory* of what needs to be computed by the cognitive system, rather than an *algorithmic-level theory* that specifies the detailed nature and time-course of the processing steps that occur in card selection. Although information gain theory offers a compelling account of many selection-task results, it will only be able to accommodate inspection-time findings by specifying a process model of selection-task behaviour.

The next theory of selection-task performance to consider is the recent instantiation of the mental models account (Johnson-Laird & Byrne, 2002). This theory proposes that with indicative versions of the selection task people will think about each of the presented cards in turn, and then select only those cards that could impact on the rule's truth or falsity. This account would seem to predict equivalent inspection times for all cards, as reasoners attempt to integrate each card with their model of the conditional to determine the impact on the rule's truth or falsity.

One important aspect of the data from Experiments 1 to 3 that warrants some further consideration in relation to the heuristic-analytic theory is the fact that participants generally were observed to spend only a very short amount of time inspecting the cards, regardless of whether they went on to select them or not. If the inspection-time data are examined across the experiments it can be seen that inspection times ranged from 1.54 s to 3.43 s for selected cards and from 0.58 s to 1.84 s for non-selected cards (transformed data in original units in all cases). Roberts and Newton (2001) also report a small inspection-time effect, finding the difference between selected and rejected cards was just 0.30 s. There would appear to be two key explanations for this smaller than expected inspection-time effect. One is that rationalisation processes are occurring very quickly, rather than being slow in nature; the other is that rationalisation processes are not occurring at all (though the latter account would still require some explanation to be provided for the small but reliable inspection-time effect that clearly does exist). Further details of the specific aspects of cognitive processing that may give rise to the small inspection-time effect with the selection task will be explored in the next chapter, in the context of the subsequent series of experiments that formed part of the present thesis.

Overall, then, the eye-tracking methodology has now been established in relation to selection-task research in Experiments 1, 2 and 3, and can profitably be applied in future research to investigate the inspection-time effect without the risk of any methodological artefacts having a bearing on the results. In addition, it appears that Evans' (e.g., 1996, 2006) heuristic-analytic theory of selection-task performance may provide a better account of card inspection-time data than other contemporary theories of the selection task. However, the *small* inspection-time effect that was produced across Experiments 1 to 3—which has also been reported elsewhere in the literature (Roberts & Newton, 2001)—certainly warrants further investigation. It is this feature of the inspection-time effect that motivated the four experiments that are reported in Chapter 3.

Chapter 3

Experiment 4 to 7

Although the experiments reported in Chapter 2 produced good evidence for a reliable inspection-time effect—with selected cards being considered for a longer time than rejected cards—the actual magnitude of this effect remained small in all three experiments. For example, across Experiments 1 to 3, mean by-participants inspection times for selected cards ranged from 1.54 s to 3.43 s, and for non-selected cards from 0.58 s to 1.84 s (transformed data in original units in all cases). Thus, selected cards were only being inspected for, at best, a second or so longer than non-selected ones. On first sight, then, the relatively small magnitude of the inspection-time effect seems to be inconsistent with Evans' (e.g., 1996, 2006) view that the effect may be attributable to the functioning of analytic *rationalisation* processes that are applied to to-be-selected cards in order to justify heuristically-determined choices. Surely, rationalisation processes should take more than a second or so to apply, since establishing a satisficing-based analytic justification for a to-be-selected card should presumably be a non-trivial operation?

There seem to be two main explanations for this apparent anomaly of a small inspection-time effect. First, the notion that a satisficing-driven rationalisation process should take a relatively long time to execute may well be misconceived. Research by Wason and Evans (1975), for example, explored the justifications that people provided for card selections, and revealed a phenomenon that they dubbed *secondary matching bias*. This is the tendency for participants to explain card selections in terms of the *matching* values that might be present on the *reverse* sides of cards. For example, given the conditional rule '*If there is an A on one side of the card then there is a 3 on the other side*', it is common for participants to give an explanation of the selection of the 'A' card by stating that a '3' on the other side (i.e., a matching value) would 'verify' the rule. Whilst this rationalisation process is conscious, evaluative and analytic, it may, however, be guided by the rapid, heuristic cueing of information.

There is no a priori reason to assume, then, that the analytic rationalisation of to-be-selected cards should take longer than a few seconds to achieve, if it is accepted that this rationalisation process can be guided by secondary matching heuristics.

There is, however, an alternative—and at least equally plausible—explanation for the small magnitude of inspection-time effect observed in Chapter 2. This explanation revolves around the idea that the analytic rationalisation of to-be-selected cards does indeed take place relatively slowly, but occurs primarily when participants *re-inspect* the rule presented to them, rather than when they inspect the actual cards. Instances of rule re-inspection occurred very frequently in all three eye-movement tracking experiments reported in Chapter 2. For example, the average frequency of rule re-inspection for Experiment 1 across all four problems was 7.50 (SD = 3.14), for Experiment 2 was 8.00 (SD = 4.97) and for Experiment 3 was 7.70 (SD = 3.01). These data can be broken down further in terms of re-inspection frequencies for each rule type within each experiment, as depicted in Table * below:

Table *: Mean rule re-inspection frequencies across all four rules in Experiments 1-3.

	Experiment 1 Mean (SD)	**Experiment 2 Mean (SD)**	**Experiment 3 Mean (SD)**	**Total mean**
If p then q	6.00 (1.87)	5.00 (2.00)	5.80 (1.30)	**5.60**
If p then not q	5.60 (1.95)	6.80 (5.45)	10.20 (3.96)	**7.53**
If not p then q	9.00 (3.39)	9.40 (6.47)	6.80 (1.64)	**7.87**
If not p then not q	9.40 (3.65)	10.80 (4.09)	8.00 (3.08)	**9.47**

Table * shows that overall, the affirmative rule (*If p then q*) receives the lowest number of re-inspections and that the double negation rule (*If not p then not q*) receives the highest number of re-inspections. The single-negation rules (*If p then not q* and *If not p then q*) lie in-between the extremes of the affirmative and double negation rules in terms of the number of re-inspections that they receive. These data confirm that people are re-inspecting rules a number of times within this presentation paradigm. More importantly, such evidence is compatible with the possibility that such rule re-inspection may be associated with time spent rationalising card choices.

3.1 Experiment 4

In an attempt to arbitrate between the two aforementioned explanations for the small magnitude of the inspection-time effect in the selection task, a new eye-movement tracking experiment was conducted—again using the abstract task with the full negations paradigm. This experiment set out to separate temporally the rule presentation from the card presentation for each task, so that participants were unable to spend time thinking about card selections whilst re-inspecting the associated conditional rule. It was predicted that if rationalisation processes are normally associated with rule re-inspection, then the visual absence of the rule might serve to shift the rationalisation process on to the to-be-selected cards themselves—hence increasing the magnitude of the inspection-time effect from that observed in Experiments 1 to 3. On the other hand, if similar inspection-time magnitudes were observed in this new experiment to those seen in Experiments 1 to 3, then this would suggest that analytic rationalisation processes indeed operate rapidly, perhaps guided by the automatic heuristic cueing of matching values for hidden sides of cards.

3.1.1 Method

Design

The design was a correlational one in which two measures were taken: whether or not a card was selected, and the associated inspection time for the card.

Participants

Participants were 30 undergraduates at the University of Derby who gained course credit for their involvement. Participants had not had any teaching on the psychology of reasoning.

Materials and apparatus

All materials and apparatus were identical to those used in Experiment 3, except that for each selection task the rule presentation was separated from the presentation of the associated cards (i.e., the rule was first presented on one sheet and followed the 2 x 2 arrangement of cards on another sheet).

Procedure

Apart from the use of a rule-separation technique, the basic experimental procedure in this experiment was identical to that adopted in Experiment 3 (i.e., a forced-choice 'yes/no' decision was required for all presented cards). The following instructions reflect the small changes to the procedure that were necessitated by the employment of the rule-separation method:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single figure number on the other side. Naturally, only one side of each card will be visible to you.
>
> For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. If you decide that a card needs to be turned over then please point to the 'yes' box under the card. If you decide that a card doesn't need to be turned over then point to the 'no' box. You will need to make a decision about each card presented to you.
>
> Please note that for each problem the rule will be presented on a separate sheet of paper to the associated cards. This means that you will need to commit each rule to memory before you tackle each problem. You have 30 seconds to do this.

As in Experiments 1 to 3, the experimenter went through the instructions with the participants to ensure that they understood them. Again, each task was presented separately and responses were made using a 20 cm metal pointer.

3.1.2 Results

Coding the inspection-time data

The same procedure was used for coding the data as in Experiment 3. Coding was terminated when a participant had finished making their decisions about all four cards on a task (i.e., when the pointer was moved away from the final card for which a 'yes/no' decision had been made).

Statistical analysis

Card selection frequencies

Before the inspection-time data were examined, standard statistical analyses were performed (as described by Evans, Ball, & Brook, 1987) to assess the pattern of card selections, and to see whether the data conformed to the response findings typically observed in the literature. A one-tailed Wilcoxon test showed that more antecedent-matching cards were selected than antecedent-mismatching ones ($z = -3.81, p < .001$) and that more consequent-matching cards were selected than consequent-mismatching ones ($z = -3.99, p < .001$). A Friedman test indicated a significant difference for the frequency of choices across logical cases ($\chi = 24.46$, df = 3, $p < .001$). As the literature suggests, there was a large preference for TA choices over all other cases and FA was the least popular choice. There was also a preference, although smaller, for TC over FC, which, as suggested in Chapter 2, could be evidence for a verification bias as well as a matching bias. Overall, then, the card-selection data revealed a typical response pattern to that seen in the selection-task literature.

Card inspection times

Selection frequencies and mean inspection times (before and after transformation) for all cards are shown in Table 3.1. The first analysis tested Prediction 1: that cards with higher selection frequencies will have higher mean inspection times. The results indicated a strong association between selection frequencies and inspection times: $r = .55, N = 16, p = .014$ (transformed data).

TABLE 3.1

Percentage frequency of card selections and overall mean inspection times (in seconds) and log times (in seconds) for each item in Experiment 4 ($N = 30$).

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	100.0		3.3		76.7		0.0	
	Natural data: time	2.49	1.68	2.02	1.51	2.55	1.43	1.81	0.94
	Transformed data: log time	0.36	0.26	0.26	0.30	0.39	0.22	0.26	0.21
	Transformed data: corrected time	2.09		1.62		2.25		1.62	
2. If p then not q	Frequency of selections	73.3		26.7		23.3		53.3	
	Natural data: time	3.46	2.38	2.26	1.48	2.47	1.66	3.99	2.67
	Transformed data: log time	0.47	0.33	0.32	0.25	0.36	0.25	0.53	0.29
	Transformed data: corrected time	2.75		1.89		2.09		3.19	
3. If not p then q	Frequency of selections	56.7		43.3		76.7		40.0	
	Natural data: time	3.14	3.81	4.89	4.34	4.00	2.62	2.85	1.85
	Transformed data: log time	0.36	0.38	0.56	0.37	0.53	0.30	0.41	0.26
	Transformed data: corrected time	2.09		3.43		3.19		2.37	
4. If not p then not q	Frequency of selections	36.7		60.0		33.3		56.7	
	Natural data: time	3.16	3.07	3.67	3.41	2.72	2.19	3.69	3.14
	Transformed data: log time	0.39	0.35	0.46	0.34	0.37	0.29	0.47	0.32
	Transformed data: corrected time	2.25		2.68		2.14		2.75	
Mean		66.68		33.3		52.5		37.5	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

The second analysis tested Prediction 2: that for any given card, mean inspection times for individuals selecting it will be greater than for those not selecting it. Table 3.2(a) shows the mean inspection times for selected and non-selected cards for all 16 cases. For two of the cards, 100% selection or non-selection rates were recorded. This meant that comparing mean inspection times for selections and non-selections was impossible for these cards. For the remaining 14 cards the difference between mean times on selected and non-selected cards (on transformed data) was in the expected direction for 6 of the 14 cards (correlations ranged from -0.34 to 0.43). This was not significant on the binomial test, indicating a failure to support Prediction 2.

The third analysis involved a test of the more powerful Prediction 3 proposed by Roberts (1998): that for each participant, mean inspection times should be longer for the cards that they have selected than for those that they have not selected. For each individual, two means were calculated from the transformed data, and a within-participants analysis of variance was performed [see Table 3.2(b)]. This confirmed Prediction 3; inspection times were longer for the cards that were selected than those that were not, $F(1, 29) = 5.64$, $MSE = 0.06$, $p = .024$.

Finally, two separate planned contrasts were carried out in order to assess the reliability of the selected versus non-selected difference for matching cards (2.75 s vs. 2.43 s; transformed data in natural units), and for mismatching cards (2.26 s vs. 1.84 s; transformed data in natural units). Both contrasts were reliable: $F(1,17) = 6.36$, $MSE = 0.22$, $p = .022$, and $F(1,17) = 19.27$, $MSE = 0.43$, $p < .001$, respectively.

TABLE 3.2

Mean inspection times for selected and non-selected cards, for Experiment 4, showing (a) inspection times (in seconds) for each item, and b) inspection times (in seconds) and log times (in seconds) overall by participants ($N = 30$).

a) by items

		Selected			*Not Selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	30	2.49	2.09	0	-	-
	FA	1	6.36	6.41	29	1.87	1.54
	TC	23	2.45	2.14	7	2.89	2.55
	FC	0	-	-	30	1.81	1.62
2. If p then not q	TA	22	3.43	2.55	8	3.53	3.19
	FA	8	3.37	2.97	22	1.86	1.62
	TC	7	2.13	1.94	23	2.58	2.09
	FC	16	4.42	3.78	14	3.50	2.68
3. If not p then q	TA	17	4.27	2.82	13	2.28	1.66
	FA	23	5.40	3.78	17	4.50	3.19
	TC	23	3.97	3.19	7	4.10	3.43
	FC	12	2.24	1.84	18	3.26	2.82
4. If not p then not q	TA	11	2.60	2.14	19	3.48	2.31
	FA	18	4.24	3.19	12	2.82	2.04
	TC	10	2.32	1.66	20	2.92	2.43
	FC	17	2.87	2.20	13	4.76	3.69

b) by participants

	Selected		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data	3.31	1.49	2.92	1.64
Transformed data: log time	0.44	0.20	0.38	0.16
Transformed data: corrected time	2.55		2.20	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

3.1.3 Discussion

Overall, the inspection-time analyses for Experiment 4 provided good support for the predictions of Evans' (e.g., 1996, 2006) heuristic-analytic account of selection-task performance, although it must be acknowledged that one of the heuristic-analytic predictions (i.e., Prediction 2) failed to show a reliable effect. This may, however, be a consequence of the lack of power associated with the by-items analysis for this prediction.

The main rationale for running Experiment 4 was to begin to investigate the issue of the relatively small magnitude of the inspection-time effect observed in Experiments 1 to 3 (as well as in Roberts and Newton's, 2001 study). One possible hypothesis was that people might be engaging in analytic rationalisation processes as they re-inspect a presented rule (rather than when looking at the to-be-selected cards themselves). By separating the presentation of the rule and the presentation of the cards an increase in the inspection-time effect for the actual cards might be found, which would support a rule re-inspection hypothesis. The alternative hypothesis was that analytic rationalisation does take place almost *entirely* during inspection of to-be-selected cards, but that this rationalisation happens very quickly.

In relation to these alternative hypotheses the present data seem definitive [Table 3.2(b)]. The mean inspection-time difference for selected cards versus non-selected cards was 0.35 s. Although this difference in inspection times was reliable, it was still small, indicating that people attended to to-be-selected items for only about a third of a second more than to-be-rejected items. This does not lend any support to the rule re-inspection hypothesis (i.e., that people are engaging in rationalisation processes whilst looking back at the conditional rule), otherwise an increased inspection-time effect should have arisen when the rule was absent from the display. The results instead suggests then that analytic rationalisation is a fairly rapid process on abstract selection-task problems. It is the detailed nature of these apparently rapid rationalisation processes that the remaining experiments in this thesis will go on to investigate.

3.2 Experiment 5

In investigating the reliable inspection-time effect that supports key tenets of Evans' (e.g. 1996, 2006) heuristic-analytic account of the selection task, Experiment 4 provided evidence to suggest that the rationalisation process that produces the inspection-time effect occurs extremely quickly. Such rapid deployment of a conscious, analytic process to justify heuristically-determined choices may, at first sight, seem anomalous. As was noted previously, it might be expected that such rationalisation should instead demand quite a slow and ponderous mode of thinking. However, we also mooted previously the idea that rapid rationalisation may not be such a strange notion after all if it is assumed that the rationalisation process is, itself, driven by the heuristically-cued consideration of possible values that might appear on the hidden sides of cards (i.e., through the operation of *secondary matching bias*; see Wason & Evans, 1975). Indeed, these notions fall directly out of the most recent version of Evans' heuristic-analytic theory (Evans, 2006), where it is argued that most people (except those of very high cognitive ability) will treat verification and falsification on the abstract selection task as though they are symmetrical, and so will happily satisfice when they deploy their analytic system, simply justifying a choice of a matching facing and hidden combination on the grounds that this will prove the rule true or prove it false (e.g., in the case of the False Consequent matching card on the negated-consequent rule form).

To investigate the nature of the analytic processing that is occurring on the selection task a method is required that allows the experimenter to have direct access to the processes as they occur. One respected method that allows this to be done is the use of verbal 'think aloud' protocols. As discussed in Chapter 2, this method has been used extensively in the reasoning and problem solving literatures and has been particularly popular since the publication of Ericsson and Simon's (e.g., 1980, 1993) research assessing the validity of the approach. Ericsson and Simon's (1993) review of a wide range of studies that had used verbal protocol techniques concluded that the elicitation of concurrent think-aloud reports from a participant engaged in task performance can provide a highly accurate and complete index of the current contents of short-term memory, in that whatever is consciously attended to by a participant is also verbalisable.

As discussed in Chapter 2, however, there is a general lack of published experiments that have utilised the verbal protocol method with abstract, indicative versions of the selection task. One study by Goodwin and Wason (1972), using only affirmative conditionals, showed that an insight into the task was demonstrated when people were asked to provide a written justification of their answers. In particular, when people selected the correct cards, they talked about falsifying the conditional whereas when people produced the typical errors that occur of the task people produced a verification explanation. In a related experiment, Wason and Evans (1975) used both affirmative (*if p then q*) and negative (*if p then not q*) statements and found that whilst people could provide insightful solutions for the negative rule, such insight was absent for the affirmative rule. Wason and Evans proposed that card choices were unconsciously determined by matching processes and then rationalised by a separate conscious and verbal process.

More recently, Evans (1995, Experiment 5) presented a protocol analysis of the indicative selection task. Protocols were scored for references to the *facing* sides of cards as well as for references to the *hidden* sides of cards. Both scores were calculated separately for selected versus rejected cards. Evans' analysis indicated that: (1) participants referred more often to the facing sides of the cards that were selected than to the facing sides of cards that ended up being rejected; and (2) participants referred more often to the hidden sides of the cards that were selected than the hidden sides of cards that ended up being rejected. These findings were consistent with Beattie and Baron's (1988) results from an earlier protocol study. Evans (1995) suggested that his results lend support to the heuristic-analytic account; people only think about some of the cards and not others, and thinking about hidden sides of cards mostly serves to rationalise decisions to choose such cards.

Although Evans' (1995) protocol-based support for the role of relevance effects and rationalisation processes in the selection task appear persuasive there are a number of important factors that should be noted. Evans' findings derive from the analysis of four distinct experimental conditions that involved very small sample sizes (i.e., *N*s of 3, 3, 4, and 5). As well as this, Evans used arbitrarily thematic selection-task materials (as opposed to purely abstract problem content), and certain experimental conditions entailed highly non-standard *judgement* instructions. Additionally, Evans' statistical analysis of his dataset using chi-squared tests was potentially problematic in that participants contributed multiple data points to both the selected and non-selected cells of the contingency tables. Finally, Evans' analyses did not focus on the important issue of the *content* of people's references to potential values that may reside on the hidden sides of cards.

It is possible, then, that gaining an understanding of whether secondary matching bias effects (Wason & Evans, 1975) are associated with hidden-side references would be especially valuable for an insight into why the processing of selected cards evidenced in inspection-time studies seems to be increased only to a small (though reliable) degree relative to non-selected cards. There seems to be clear scope for replicating Evans' (1995) protocol-based findings with an increased sample size, more conventional task features—including the employment of standard abstract problems—and traditional task instructions. Pursuing such a replication was the primary aim of Experiment 5. The heuristic-analytic predictions that have been applied effectively in previous inspection-time studies were adapted (e.g., Evans, 1996; Experiments 1-4 of the present thesis; Roberts, 1998b) so as to enable more powerful statistical tests to be pursued of the heuristic-analytic theory in terms of people's references to the *facing* and *hidden* sides of selected and non-selected cards. Three key predictions were therefore established:

Prediction 1: Cards that are associated with higher selection rates will also be associated with more references to their facing sides.

Prediction 2: For any given card, those participants who select it will refer more to its facing side than those participants who do not select it.

Prediction 3: For each participant, their mean number of references to the facing sides of selected cards should be higher than to the facing sides of non-selected cards.

The latter participant-level prediction is a version of that promoted by Roberts (1998b) in the context of card inspection-time analyses, and is argued to be a more powerful test of the heuristic-analytic account than either Prediction 1 or Prediction 2, which involve item-level analyses. All three predictions have been stated solely in terms of references to the *facing* sides of cards. It is also possible, however, to restate each of these predictions so that they apply equally to the analysis of references to the *hidden* sides of cards. Such re-stated predictions would be entirely in line with the claim of the heuristic-analytic theory (e.g., Evans 1995, 1996, 2006) that rationalisation processes serve merely to justify card choices, thereby promoting increased references to hidden sides of selected cards relative to the hidden sides of non-selected ones. Both sets of predictions were adopted for the purpose of the present experiment.

Finally, one further prediction was derived from Evans' (1996, 2006) heuristic-analytic theory that is associated to the content of people's explicit references to the hidden sides of cards. This prediction was as follows:

Prediction 4: The total pool of references to hidden sides of cards should be dominated by references to potential matching values that might appear on the reverse sides of cards relative to either mismatching values or negated matching values.

This last prediction derives from the assumption that secondary matching heuristics may guide the analytic rationalisation processes associated with to-be-selected cards (Wason & Evans, 1975).

3.2.1 Method

Participants.

Participants were 30 undergraduate volunteers from the University of Derby who took part in the experiment to gain course credit. Participants had not received any tuition on the psychology of reasoning.

Materials and apparatus.

The experiment involved selection tasks employing the same abstract conditional rules within the standard negations paradigm as in Experiments 1 to 4. Again, each participant received four versions of the task. Each problem was presented on a single A4 page. The rule was positioned at the top of the page, a reminder of the task requirement appeared in the middle of the page, and the pictures of the four cards were presented in the lower half of the page in a 2 x 2 arrangement. The location of cards within each array was always random. The experiment was carried out in an audio-recording suite to enable participants' think-aloud protocols to be recorded.

Procedure

Participants were tested individually. They were initially told about the essential nature of the experiment and an explanation was provided of the 'think-aloud' requirement. To help clarify the expectations surrounding the think-aloud procedure and to put participants at their ease, a brief, video-based demonstration was provided of someone verbalising whilst carrying out a moderately difficult problem-solving task involving the rebuilding of a pyramid structure using jigsaw-like building blocks. Subsequent to this demonstration the following written instructions were presented:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. These problems will appear on separate sheets in front of you. Each problem consists of four

cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single-figure number on the other side. Naturally only one side of each card will be visible to you.

For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. It is all right for you to change your mind as you work through a problem, and I will not record any decisions until you tell me what your final choice or choices are.

Whilst you are reading through each problem and deciding how to solve it, please remember that I would like you to think aloud. As I've explained, you should find it quite natural to say aloud whatever happens to come into your head whilst you are working on these tasks. If you do fall silent for any length of time, however, I will gently prompt you to try and keep thinking aloud.

Once the participant had read the instructions the experimenter re-read them aloud and provided an opportunity for participants to seek clarification concerning any of the study requirements. The four problems were then presented in a random order.

3.2.2 Results

Protocol coding, reliability assessment, and normality checks

Before inferential analysis, the verbal protocols were transcribed and then coded using three categorisation systems. The first system, inspired by Evans (1995, Experiment 5), involved examining each participant's protocol and, for each rule, identifying their unique *references to the facing sides* of each of the four presented cards. Frequency counts of the number of references per card were then calculated and provided a measure for use in subsequent statistical analyses. In applying this first scheme any references to facing card sides that occurred when participants were making or confirming their final card selections were not coded. This would avoid the possibility of obtaining artefactual support for heuristic-analytic predictions arising from the fact that only selected cards needed to be actively registered by participants. Any references to facing sides arising at the selection-registering phase would artificially inflate the frequency-count of mentions to selected cards, since it is only these cards that need to be referred to explicitly. This would have been a similar problem as identified in Experiment 1 in terms of a 'pointing bias' that could have arisen when registering card selections (which could have inflated inspection times for such cards). This conservative measure of references to facing sides provides a stronger test of heuristic-analytic predictions than the coding scheme applied by Evans (1995, Experiment 5), which appears not to have considered such methodological artefacts. Frequency counts for each participant's total number of references per card were then calculated in order to provide a measure for use in subsequent statistical analyses.

The second categorisation system was identical to the previous one in all respects, except for its focus on participants' *references to hidden sides* of each of the presented cards. Two coders independently applied both of the aforementioned categorisation schemes to the full set of verbal protocols. Inter-coder reliability checks revealed a very high degree of consistency between coders (i.e., 97% inter-coder agreement), and there was no evidence of systematic divergences between coders in their categorisation of discrete references to the facing or hidden sides of each logical case. The codes applied by the thesis author were used for all subsequent analyses associated with the experimental predictions.

The third categorisation system involved sub-categorising each reference to a hidden side in terms of the specific letter or number content mentioned in that reference. This coding scheme used the following four sub-categories, which are illustrated in terms of participants' potential references to what might have been on the other side of the A card associated with an *'If A then 3'* rule: (1) a reference to a *matching* item (e.g., mentioning the possibility of a '3' on the other side of the 'A' card); (2) a reference to *mismatching* item (e.g., mentioning the possibility of a number such as a '7', on the other side of the 'A'); (3) a reference to a *negated matching* item (e.g., stating that there could be a number that is 'not a 3' on the reverse of the 'A'); and (4) a *non-*

specific reference to what might be on the other side of the card (e.g., when participants stated how 'It is important to see what's on the other side of the A', without qualifying such a comment further). It should be noted that whilst other sub-categories are possible in addition to the four described here (e.g., references to negated mismatching items), the present four sub-categories successfully captured the full range of content that was distinguished in participants' references to the hidden sides of cards. As there was only limited scope for miscategorising references using this scheme (i.e., the new codes simply reflected a more detailed breakdown of the explicit references to hidden card sides that had already been identified) it was not necessary to pursue inter-coder reliability checks on the application of these codes.

Statistical analysis

Descriptive analysis of the data revealed that they were positively skewed. Log transformations applying a constant of 0.4 was used for both the facing-side and the hidden-side data transformations and were found to stabilise variances successfully. For clarity of interpretation the results section reports means both before transformation and converted back into their original units after transformation.

Card selection frequencies

The first concern was to assess whether the four selection tasks elicited the standard pattern of card selections observed in the literature (i.e., more matching than mismatching choices across antecedent and consequent cases). Matching bias was examined once again using the procedures adopted by Evans, Ball, and Brooks (1987). Wilcoxon signed-rank tests (one-tailed) revealed that more antecedent-matching cards were selected than antecedent-mismatching ones ($p = .01$), and that more consequent-matching cards were selected than consequent-mismatching ones ($p < .001$). This pattern of results is, therefore, typical of that seen for selection tasks within the negations paradigm.

TABLE 3.3

Percentage frequency of card selections and overall mean references to facing and hidden sides for each item in Experiment 5 ($N = 30$).

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	90.0		16.7		60.0		10.0	
	Facing Side - ND	1.37	0.85	0.57	0.94	0.90	0.71	0.53	0.90
	Facing Side - TD	1.19		0.29		0.67		0.28	
	Hidden Side - ND	0.87	0.78	0.17	0.46	0.53	0.51	0.17	0.38
	Hidden Side - TD	0.67		0.08		0.38		0.09	0.67
2. If p then not q	Frequency of selections	86.7		10.0		10.0		63.3	
	Facing Side - ND	1.53	1.14	0.37	0.67	0.37	0.67	0.80	0.85
	Facing Side - TD	1.22		1.30		1.30		0.53	
	Hidden Side - ND	1.13	0.86	0.07	0.25	0.10	0.31	0.53	0.63
	Hidden Side - TD	0.89		0.04		0.06		0.34	
3. If not p then q	Frequency of selections	60.0		43.3		66.7		43.3	
	Facing Side - ND	1.30	1.29	1.07	1.08	1.17	1.26	0.90	1.27
	Facing Side - TD	0.86		0.70		0.75		0.49	
	Hidden Side - ND	0.87	1.07	0.83	1.21	0.73	1.02	0.47	0.86
	Hidden Side - TD	0.56		0.43		0.39		0.23	
4. If not p then not q	Frequency of selections	60.0		46.7		33.3		53.3	
	Facing Side - ND	1.13	1.14	1.10	0.92	0.70	0.84	1.10	1.03
	Facing Side - TD	0.75		0.78		0.43		0.75	
	Hidden Side - ND	0.47	0.86	0.60	0.72	0.20	0.41	0.63	0.72
	Hidden Side - TD	0.25		0.38		0.11		0.39	
Mean		74.2		29.2		25.0		42.5	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent, ND = natural data.; TD = transformed data (in original units).

Verbal protocol analyses

The statistical analyses examined the four predictions, identified above, that derive from Evans' (e.g., 1996, 2006) heuristic-analytic account. As noted previously, Prediction 1, Prediction 2 and Prediction 3 apply equivalently to the measure of references to the facing sides of cards and to the measure of references to the hidden sides of cards, so all predictions have two sets of analysis: one for the facing sides and one for the hidden sides.

The first analysis tested Prediction 1: Cards associated with higher selection rates will be associated with more references to their facing (and hidden) sides. This analysis involved exploring the correlation across all 16 cards between the overall mean references to a card side and the card's associated selection frequency (refer to Table 3.3). The correlation for the facing sides showed that there was a strong positive association between the mean number of references to facing sides and card selection frequencies, $r = .94$, $N = 16$, $p < .001$ (transformed data). The correlation between selection frequency and mean references to hidden sides was also significant, $r = .89$, $N = 16$, $p < .001$ (transformed data).

The second analysis tested Prediction 2: that for any given card, mean references to a card side for individuals selecting it would be higher than for those who did not. Assessing Prediction 2 involves determining, for each card, whether the mean reference to a card side is higher for selectors than non-selectors. Mean references to facing and hidden sides for each card are given in Table 3.4. After transformation, mean references to facing sides for 16 out of 16 cards were greater for selectors than non-selectors ($p < .001$), two-tailed with the binomial test. Mean references to hidden sides for 15 out of 16 cards were greater for selectors than non-selectors cases, significant with the binomial test ($p = .001$, two-tailed).

TABLE 3.4

Mean number of references to facing and hidden sides (by items) for selected and non-selected cards, for Experiment 5 ($N = 30$).

		Selected					*Non-selected*				
			Facing		*Hidden*			*Facing*		*Hidden*	
Rule	*Card*	*N*	*TD*	*ND*	*TD*	*ND*	*N*	*TD*	*ND*	*TD*	*ND*
1. If p then q	TA	27	1.30	1.44	0.72	0.93	3	0.51	0.67	0.20	0.33
	FA	5	1.11	1.60	0.53	0.80	25	0.19	0.36	0.02	0.04
	TC	18	1.26	1.33	0.67	0.78	12	0.15	0.25	0.09	1.67
	FC	3	1.64	2.00	0.51	0.67	27	0.20	0.37	0.06	0.11
2. If p then not q	TA	26	1.34	1.54	0.89	1.12	4	0.75	1.50	0.89	1.25
	FA	3	2.00	2.00	0.51	0.67	27	0.10	0.19	-0.01	0.00
	TC	3	2.00	2.00	1.01	1.00	27	0.10	0.19	-0.01	0.00
	FC	19	0.89	1.05	0.58	0.74	11	0.15	0.36	0.07	0.18
3. If not p then q	TA	18	1.74	2.00	1.08	1.33	12	0.15	0.25	0.09	1.67
	FA	13	1.42	1.54	0.89	1.31	17	0.32	0.71	0.20	0.47
	TC	20	1.01	1.40	0.53	0.90	10	0.36	0.70	0.18	0.40
	FC	13	1.15	1.62	0.75	1.08	17	0.18	0.35	-0.01	0.00
4. If not p then not q	TA	18	1.05	1.50	0.41	0.72	12	0.38	0.58	0.05	0.08
	FA	14	1.38	1.50	0.80	1.00	16	0.41	0.75	0.13	0.25
	TC	10	0.92	1.20	0.26	0.40	20	0.26	0.45	0.06	0.10
	FC	16	1.51	1.69	0.86	1.06	14	0.25	0.43	0.08	0.14

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent, ND = natural data; TD = transformed data (in original units).

The third prediction was Roberts (1998) participant-level analysis. It tested that for each individual, the mean references to sides of cards should be higher for selected than for non-selected cards. Two mean references to facing side scores and hidden side scores were calculated for each person from the transformed data [Table 3.5]. A within-participants analysis of variance (ANOVA) provided good support for Prediction 3 for both facing sides, $F(1, 29) = 115.44$, $MSE = 2.54$, $p < .001$, and for hidden sides, $F(1, 29) = 106.43$, $MSE = 2.13$, $p < .001$.

TABLE 3.5

Mean number of references to facing and hidden sides (by participants) for selected versus non-selected cards for Experiment 5.

	Selected				*Non-selected*			
	Facing		*Hidden*		*Facing*		*Hidden*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
ND	1.52	0.51	0.98	0.49	0.46	0.49	0.16	0.26
TD (log)	0.22	0.13	0.05	0.19	-0.19	0.22	-0.32	0.12
TD	1.26		0.73		0.24		0.07	

Note: ND = natural data; TD (log) = transformed data (in $\log^{10}$ units); TD = transformed data (in original units).

Secondary matching bias predictions were assessed pertaining to Prediction 4 by taking the total pool of references to *hidden* sides of cards produced by all 30 participants, and then computing the distribution of references within this pool across the four sub-categories of reference-type (i.e., matching items, mismatching items, negated matching items, and non-specific references). This analysis revealed that the mention of matching values dominated people's verbalisations concerning what might appear on the reverse sides of cards (64% of references) in relation to the mention of negated matching values (35% of references), mismatching values (< 1% of references) and unspecified values (< 1% of references).

3.2.3 Discussion

The predictions that were derived from Evans' (e.g., 1996, 2006) heuristic-analytic account of the selection task were supported in Experiment 5. The results demonstrate that selected cards are associated with more attention than non-selected cards, which can be seen by the number of explicit references to both their facing sides and their hidden sides. This particular finding in relation to the hidden sides of the cards appear to lend support to Evans' claimed role for analytic rationalisation processes during selection-task performance. So although the to-be-selected cards encourage participants to consider what values might be on their reverse sides, this consideration does *not* seem to the change the fact that these cards tend to end up selected (otherwise the link between references to hidden sides and card selection would be broken). One interpretation of the findings, then, is that thinking about the hidden sides of cards appears to have a minimal functional role in determining card choices (at least for a substantial number of participants). Instead it seems that thinking about the hidden sides of the cards mainly serves to confirm decisions to go ahead and choose such cards (cf. Evans, 1995, p. 168).

The experiment also addresses the issue of what people are actually thinking about when they consider the reverse sides of cards. The findings are clear-cut. First, participants do not think at all about potential mismatching values that may appear on the reverse sides of cards. This may be taken as further support for the heuristic-analytic view that people tend not to see mismatching values as having any relevance to their decision-making during the evaluation of conditional statements. Second, the finding that people's consideration of hidden values is dominated by matching possibilities seems to be in line with Wason and Evans' (1975) notion that secondary matching heuristics may cue people's analytic accounts as to why values on the hidden sides of cards justify selection of those cards. This evidence for secondary matching effects in abstract selection tasks also helps make sense of card inspection-time findings (as reported in Experiments 1 to 4), which suggest that analytic rationalisation processes are rapid in nature. Rationalisation might well be expected to be extremely fast if people's justifications are facilitated by the heuristic cueing of 'relevant' (i.e., matching) values that could appear on the reverse sides of cards.

3.3 Experiment 6

The verbal protocol method in Experiment 5 was successful in a number of ways. First, the results of the study supported the predictions that can be derived from the heuristic-analytic theory of the selection task. Second, the findings are also in line with results from previous mouse-tracking and eye-tracking studies of card inspection times and from verbal protocol studies (Evans, 1995, 1996; Experiments 1 to 4 in the present thesis). Third, Experiment 5 was additionally able to clarify the important role played by secondary matching processes when people are referring to the hidden sides of to-be-selected cards.

There is, however, an interesting task-presentation issue associated with Experiment 5 that seems to warrant further empirical examination. To explain this issue, we first note that two experiments in the present research programme (Experiments 3 and 4) used a forced-choice methodology, whereby participants had to register 'select/don't select' decisions for all presented cards (see also Evans et al. 1987; Roberts, 1998b, Experiment 3). Experiment 5, however, employed a standard methodology whereby participants only had to register their 'select' decisions. All of the previous experiments that have used a forced select/reject requirement have still demonstrated that matching responses dominate selections, even though participants are required to attend and respond to cards that were normally paid little attention. Another important finding from the present research programme (Experiment 3) is that the inspection-time effect (whereby selected cards are looked at for longer than rejected cards) is also not totally undermined by the use of a 'select/don't select' decision requirement, although the effect does seem to be reduced in magnitude, presumably because rejected cards now become associated with at least some (enforced) consideration.

What is of general interest here, then, is the issues of what, exactly, people think about when they are compelled to inspect cards that they would not ordinarily attend to? In particular, do people who are making card selections within the 'select/don't select' paradigm think beyond the facing sides of those cards that they choose to reject? According to the heuristic-analytic theory people should *not* think about what might be on the hidden sides of to-be-rejected cards. The enforced decision requirement would mean that people would have to attend to such cards as they are required to make an active 'don't select' response to them, but the fact that these cards should be rapidly deemed *irrelevant* means that analytic rationalisation processes would not be called upon (cf. Evans 1998a). In summary, rationalisation processes in the selection task (and perhaps more generally too) are assumed to be asymmetrical, in that people only pursue analytic justifications for cards that they wish to select (as cued by relevance), but not for cards which they wish to reject (on the basis of perceived irrelevance).

Experiment 6 was undertaken to test these heuristic-analytic predictions in relation to the enforced-decision paradigm by, once again, using verbal protocol analysis. The experiment used the same abstract selection tasks and think-aloud instructions employed in Experiment 5, except for the presence of enforced 'select/don't select' requirements for all cards. In order to test fully the heuristic-analytic theory, the equivalent set of predictions for both facing and hidden card sides as used in Experiment 5, was employed. It was expected that there would be some possible weakening of effect sizes for the *facing side* predictions (Prediction 1, Prediction 2, and Prediction 3) owing to the enforced decision procedure. However, the previous inspection-time data from Experiment 3, led to the expectation that the basic finding of increased attention to selected cards over rejected ones would remain intact (i.e., people would give to-be-rejected cards only a minimal amount of explicit consideration, dwelling instead on to-be-selected cards). In terms of predictions for the hidden sides of cards, effects of broadly similar magnitude to those that arose in Experiment 5 for Prediction 1 to Prediction 4 were anticipated (i.e., participants were not expected to think about the reverse sides of to-be-rejected cards any more than in the standard selection-task paradigm).

3.3.1 Method

Participants

Participants were 30 undergraduate volunteers from the University of Derby, who obtained course credit for their involvement in the study. No participants had received prior tuition concerning the psychology of reasoning.

Materials and apparatus

The selection-task materials and apparatus were identical to those used in Experiment 5, except that under each card participants were now also presented with small 'yes' and 'no' decision boxes (separated horizontally from the card by 0.8 cm) at a distance of about 1 cm from its lower edge.

Procedure

The procedure was identical to Experiment 5, with the exception that instructions were modified to include reference to the presence of 'yes' and 'no' response boxes below each card. The instructions therefore read as follows:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. These problems will appear on separate sheets in front of you. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single-figure number on the other side. Naturally only one side of each card will be visible to you.
>
> For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. You will need to make a 'turn/don't turn' decision about all the cards presented to you. The 'yes' and 'no' boxes underneath each card are present to remind you that you must make a 'turn/don't turn' decision for every card. It is all right for you to change your mind as you work through a problem, and I will not record any decisions until you tell me what your final answers are for each card.
>
> Whilst you are reading through each problem and deciding how to solve it, please remember that I would like you to think aloud. As I've explained, you should find it quite natural to say aloud whatever happens to come into your head whilst you are working on these tasks. If you do fall silent for any length of time, however, I will gently prompt you to try and keep thinking aloud.

As in Experiment 5, once the participant had read the instructions the experimenter then read them aloud once more to enable any clarification to be sought concerning the task requirements. The four problems were presented in a random order.

3.3.2 Results

Protocol coding, reliability assessment, and normality checks

Transcribed protocols were coded using identical categorisation and scoring schemes as had been applied in Experiment 5. Inter-coder reliability checks revealed a high level of consistency between coders in their application of the categorization schemes pertaining to references to facing and to hidden sides of cards (i.e., 95% inter-coder agreement). The codes applied by the thesis author were used for all subsequent analyses associated with the experimental predictions (Prediction 1 to Prediction 4).

Statistical analysis

Again, descriptive analysis of the data revealed that they were positively skewed. Log transformations were applied, but this time a constant of 0.6 was used for the facing-side data transformations and 0.2 for the hidden-side data transformations. These transformations were found to stabilise variances successfully. For clarity of interpretation means are reported both before transformation and converted back into their original units after transformation.

<u>Card selection frequencies</u>

Wilcoxon signed-ranks tests (one-tailed) revealed that more antecedent-matching cards were selected than antecedent-mismatching ones, $p < .001$, and that more consequent-matching cards were selected than consequent-mismatching ones, $p < .001$. The standard matching-bias pattern is, therefore, strongly evident in the card-selection responses associated with this enforced decision paradigm.

<u>Verbal protocol analyses</u>

The mean number of references to the facing and the hidden sides of each card, and each card's overall selection frequency, are presented in Table 3.6. The correlations for P1 between mean references to card sides and selection frequencies were significant for *facing* sides, $r = .88$, $N = 16$, $p < .001$ (transformed data), and for *hidden* sides, $r = .94$, $N = 16$, $p < .001$ (transformed data).

TABLE 3.6

Percentage frequency of card selections and overall mean references to facing and hidden sides for each item in Experiment 6 ($N = 30$).

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	80.0		23.3		80.0		16.7	
	Facing Side - ND	1.57	0.82	0.93	0.74	1.43	0.77	0.90	0.71
	Facing Side - TD	1.44		0.75		1.26		0.72	
	Hidden Side - ND	0.67	0.55	0.17	0.38	0.60	0.50	0.10	0.31
	Hidden Side - TD	0.43		0.07		0.39		0.04	
2. If p then not q	Frequency of selections	86.7		33.3		33.3		70.0	
	Facing Side - ND	1.87	0.90	0.83	0.70	0.93	0.83	1.37	0.81
	Facing Side - TD	1.74		0.66		0.75		1.22	
	Hidden Side - ND	1.10	0.96	0.27	0.45	0.30	0.65	0.67	0.61
	Hidden Side - TD	0.67		0.12		0.12		0.40	
3. If not p then q	Frequency of selections	56.7		46.7		66.7		43.3	
	Facing Side - ND	1.43	1.33	1.30	0.95	1.40	1.10	0.77	0.57
	Facing Side - TD	1.18		1.06		1.14		0.63	
	Hidden Side - ND	0.90	1.24	0.53	0.68	0.67	0.92	0.37	0.56
	Hidden Side - TD	0.45		0.26		0.33		0.17	
4. If not p then not q	Frequency of selections	56.7		50.0		40.0		50.0	
	Facing Side - ND	1.17	0.75	1.27	1.05	1.10	0.76	1.00	0.95
	Facing Side - TD	1.02		0.99		0.95		0.72	
	Hidden Side - ND	0.50	0.57	0.60	0.62	0.33	0.55	0.53	0.68
	Hidden Side - TD	0.27		0.34		0.15		0.26	
Mean		70.0		38.3		55.0		45.0	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent, ND = natural data; TD = transformed data (in original units).

TABLE 3.7

Mean number of references to facing and hidden sides (by items) for selected and non-selected cards, for Experiment 6 ($N = 30$).

			Selected					*Non-selected*			
			Facing		*Hidden*			*Facing*		*Hidden*	
Rule	*Card*	*N*	*TD*	*ND*	*TD*	*ND*	*N*	*TD*	*ND*	*TD*	*ND*
1. If p then q	TA	24	1.49	1.58	0.52	0.75	6	1.31	1.50	0.16	0.33
	FA	7	0.99	1.14	0.13	0.29	23	0.69	0.87	0.05	0.13
	TC	24	1.40	1.50	0.51	0.71	6	0.95	1.17	0.07	0.17
	FC	5	1.35	1.40	0.09	0.20	25	0.63	0.80	0.03	0.08
2. If p then not q	TA	26	1.64	1.81	0.67	1.12	4	2.09	2.25	0.69	1.00
	FA	10	1.10	1.20	0.29	0.50	20	0.47	0.65	0.06	0.15
	TC	10	1.10	1.30	0.45	0.80	20	0.60	0.75	0.02	0.05
	FC	21	1.31	1.48	0.45	0.81	9	0.99	1.11	0.16	0.33
3. If not p then q	TA	17	1.22	1.29	0.48	0.76	13	1.10	1.62	0.39	1.07
	FA	14	1.14	1.36	0.40	0.71	16	0.95	1.25	0.16	0.38
	TC	20	1.22	1.35	0.39	0.60	10	0.99	1.50	0.23	0.80
	FC	13	0.63	0.77	0.20	0.39	17	0.63	0.77	0.15	0.35
4. If not p then not q	TA	17	1.31	1.17	0.46	0.71	13	0.69	0.85	0.10	0.23
	FA	15	1.35	1.53	0.61	0.87	15	0.66	1.00	0.16	0.33
	TC	12	1.14	1.17	0.31	0.58	18	0.81	1.06	0.07	0.17
	FC	15	0.72	0.93	0.36	0.67	15	0.72	1.07	0.18	0.40

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent, ND = natural data; TD = transformed data (in original units).

Mean references to sides of cards (both before and after transformation) for selections and non-selections, for each of the 16 cards, are given in Table 3.7. For transformed data, the difference between mean references to facing sides for selected and non-selected cards was in the expected direction for 13 out of 16 cases (two ties), which was significant with a binomial test, $p = .021$, two-tailed. The difference between mean references to hidden sides for selected and non-selected cards was in the expected direction for 15 out of 16 cases, which was significant with a binomial test, $p = .001$, two-tailed.

To assess Prediction 3 the more powerful participant-level analyses using ANOVA was used, which revealed (see Table 3.8) a significant difference in the mean references to *facing* sides for participants' selected versus non-selected cards, $F(1, 29) = 4.62$, $MSE = .04$, $p = .04$, and a significant difference in the mean references to *hidden* sides for participants' selected versus non-selected cards, $F(1, 29) = 8.94$, $MSE = .40$, $p = .006$.

TABLE 3.8

Mean number of references to facing and hidden sides (by participants) for selected versus non-selected cards for Experiment 6.

	Selected				*Non-selected*			
	Facing		*Hidden*		*Facing*		*Hidden*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
ND	1.33	0.47	0.67	0.47	1.10	0.48	0.37	0.27
TD (log)	0.23	0.12	-0.27	0.25	0.17	0.14	-0.43	0.20
TD	1.08		0.34		0.88		0.17	

Note: ND = natural data; TD (log) = transformed data (in $\log^{10}$ units); TD = transformed data (in original units).

Finally, secondary matching bias predictions associated with Prediction 4 were assessed by calculating the distribution of all participants' references to *hidden* sides across the four sub-categories of reference-type: matching items, mismatching items, negated matching items, and non-specific references. The mention of matching values dominated participants' comments about what might appear on the reverse sides of cards (62% of references) in relation to the mention of negated matching values (33% of references), mismatching values and unspecified values (< 3% of references in each case). This distribution of references to hidden sides across these four categories is strikingly similar to the distribution observed in Experiment 5.

3.3.3 Discussion

As with Experiment 5, the results of Experiment 6 were consistent with the predictions of the heuristic-analytic account of performance on abstract versions of the selection task. The results specifically show that people referred more to the facing and hidden sides of those cards that they ended up selecting relative to those cards that they ended up rejecting. This finding thus persists despite the use of an enforced-decision paradigm which requires people to give at least some attention to cards that they might ordinarily simply ignore on the basis of their perceived *irrelevance* to the task they are engaged in. It was also anticipated that effect sizes in relation to Predictions 1, 2 and 3 would differ slightly from those in Experiment 5. In particular, it was predicted that the magnitude of the effect size for *facing-side* predictions would weaken as participants are forced to attend to all four cards in order to register a 'select/don't-select' decision for each of them. In relation to the *hidden-side* predictions it was anticipated that there would be no real change in the magnitude of the effect size as participants were not expected to think any more about the reverse sides of the to-be-rejected cards even when they were forced to consider their facing sides.

These expectations were shown to be correct in all respects. For example, in relation to the Prediction 2, item-based analysis, whereas 16 out of 16 cards in Experiment 5 showed increased references to *facing* sides for selectors compared to non-selectors, this dropped slightly to 13 out of 16 cards in Experiment 6. In contrast, there was no such drop between Experiments 5 and 6 in terms of references to *hidden* sides for selectors compared to non-selectors across cards (i.e., 15 out of 15 cards showed expected differences in both experiments). A similar pattern of changes to effect magnitudes was seen across Experiments 5 and 6 in relation to the P3 participant-based analyses. From Tables 3.5 and 3.8 it can be seen that the *mean difference* in references to *facing* sides for selected versus non-selected cards dropped quite markedly from 1.02 references in Experiment 5 (i.e. 1.26 minus 0.24) to 0.20 references in Experiment 6 (i.e., 1.08 minus 0.88), whereas the *mean difference* in references to *hidden* sides for selected versus non-selected cards dropped less strikingly from 0.66 references (i.e., 0.73 minus 0.07) in Experiment 5 to 0.17 references (i.e., 0.34 minus 0.17) in Experiment 6.

3.4 Experiment 7

Experiments 5 and 6 aimed to employ concurrent verbal protocols to investigate heuristic-analytic predictions regarding what people think about when they are engaging in abstract versions of the selection task. Experiment 5 used a standard selection-task paradigm, where active, select decisions were only required for those cards participants felt needed to be selected. In contrast, Experiment 6 assessed the impact on the content of people's thinking of imposing an enforced 'select/don't-select' decision requirement on all four cards associated with each presented task. Overall, Experiments 5 and 6 uncovered very good protocol-based evidence to support the view that the perceived 'relevance' of information has a major influence on both the heuristic and the analytic processing that arises during abstract selection-task performance, as predicted by Evans' heuristic-analytic theory (e.g., 1996, 2006).

In particular, Experiments 5 and 6 both demonstrated that selected cards are associated with more attention than non-selected cards, which can be seen by the number of explicit references that people make to their facing sides as well as their hidden sides. In addition, thinking about the hidden sides of cards appears to have little *functional* role in determining card choices, instead serving mainly to confirm decisions to choose such cards. Finally, in relation to the important issue of why the inspection-time effect seen in Experiments 1 to 4 might be so small in magnitude, Experiments 5 and 6 provided data that support the view that any analytic rationalisation processes applied to to-be-selected cards may be rapid in nature because of the apparent role of *secondary matching* processes that cue people to think about *matching* values that may be on the reverse sides of matching cards. The finding that people's consideration of hidden values is also dominated by matching possibilities seems to be entirely in line with Evans' (2006) notion of a satisficing-oriented rationalisation process underpinning much analytic reasoning.

There does, however, seem to be yet one more issue that needs investigating in relation to these verbal-protocol experiments. A remaining concern is that it may well be that the actual act of carrying out of a concurrent think-aloud requirement interferes with the *normal* processes of reasoning that occur on the selection task. There is, in fact, no a priori reason to expect such reactivity given Ericsson and Simon's (e.g., 1993) arguments about Level 1 and Level 2 verbalisation processes being largely immune to such reactive effects on primary task-based processes. But it would seem to be better to eliminate—through empirical means—any possibility that the think-aloud requirement may have a reactive effect on the normal process of reasoning on the abstract selection task.

One way to examine the reactivity issue empirically would be to run a study that required participants to produce verbal protocols whilst *simultaneously* taking a measure of their eye-movements. If the eye-movement data were disrupted—for example, if the inspection-time effect found in Experiments 1 to 4 was eradicated, reversed, or even of far greater magnitude—then this would provide evidence for a reactive effect of the verbalisation requirement on the normal processing that occurs during selection-task performance. Any demonstration of such reactivity could seriously weaken the protocol-based evidence for the heuristic-analytic theory obtained in Experiments 5 and 6. If, however, the eye-movement-based inspection-time effect under a verbal-protocol requirement was identical or very similar to that observed in the previous eye-tracking experiments in this thesis, then the findings from Experiment 5 and 6 would seem to be validated, and the support that these findings lend to the heuristic-analytic theory would likewise appear to be sound.

Experiment 7 set out to examine the reactivity issue head on. To this end, a selection-task experiment was established that employed the full negations paradigm with an enforced-decision requirement, in addition to an instructional request for participants to provide concurrent verbal protocols during their task-based processing. Whilst these verbal protocols were being generated, participants' eye-movements could also be tracked using the same technique as in previous experiments. The eye-tracking data could subsequently be analysed to assess whether concurrent verbalisation had any effect on the nature or magnitude of the inspection-time effect.

The eye-tracking data could also provide information as to whether there was any effect of the verbalisation requirement on the *overall time taken for card selections* associated with the presented tasks. This issue is not related to reactivity effects, but is instead directly concerned with arbitrating between the possible role of Level 1 versus Level 2 verbalisation processes in protocol production with the selection task. As discussed earlier, Level 1 verbalisation is simply the vocalisation of current thoughts that are in a verbal form anyway, and it is therefore viewed as neither impacting on problem-solving times nor on the structure of the thinking process (cf. Ericsson & Simon, 1993). Evans (e.g., 1989) would argue that reasoning with the selection task is primarily verbally-based, and therefore no influence of verbalisation on the time-course (or structure) of thinking would be expected. Level 2 verbalisation, in contrast, involves vocalising current thoughts that are in a *non-verbal* form. Because this involves a re-coding process, it has been observed to *slow down* primary task performance (whilst not impacting upon the structure of the process; see Ericsson & Simon, 1993, for relevant evidence). Johnson-Laird (e.g., 1985) has suggested that reasoning involves model-based mental representations that may utilise a visuo-spatial mental substrate. As such, verbalisation that arises during reasoning might be expected to involve an element of re-coding from model-based, visuo-spatial representations to verbal output. This re-coding might well lead to a detectable effect of a Level 2 verbalisation requirement on the time-course of thinking, generally extending inspections times uniformly across all cards.

3.4.1 Method

Participants

Participants were 31 undergraduate volunteers at the University of Derby who took part in the experiment in order to gain course credit. Participants had not received any tuition on the psychology of reasoning.

Materials and apparatus

The same selection tasks employing the same abstract conditional rules within the standard negations paradigm were utilised as in Experiments 1 to 6. As before, each participant received four versions of the task. Each problem was presented on a single A4 page. The rule was positioned at the top of the page, a reminder of the task requirement appeared in the middle of the page, and the pictures of the four cards were presented in the lower half of the page in a two-by-two arrangement. The location of cards within each array was always random. Under each card participants were also presented with small 'yes' and 'no' decision boxes (separated horizontally from the card by 0.8 cm) at a distance of about 1 cm. from its lower edge. As this experiment involved the use of both eye-tracking and verbal-protocol methodologies, the problems were presented vertically on an angled table at a distance of approximately 0.6 m from the participant. The participant was seated in an adjustable chair so that their position and height in relation to the eye-tracker could be optimised. There was audio-recording equipment present within the laboratory to record the participant's verbal protocols. The eye-tracking equipment and setup were identical to those used in Experiments 1 to 4.

Procedure

Participants were tested individually. The basic nature of the experiment was explained to them in terms of the use of eye-tracking equipment and the requirement for them to produce concurrent verbal protocols. The experimenter then went through the expectations concerning the think-aloud procedure particularly carefully, with the use of the video example as employed in Experiments 5 and 6. Participants then had their eye-movements calibrated using the nine-point calibration matrix, as in Experiments 1 to 4. Once the participants point-of-gaze coordinates had been determined, participants were presented with the task instructions as follows:

> This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. These problems will appear on separate sheets in front of you. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on

one side and a single figure number on the other side. Naturally only one side of each card will be visible to you.

For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. If you decide that a card needs to be turned over then please point to the 'yes' box under the card. If you decide that a card doesn't need to be turned over then point to the 'no' box below the card. You will need to make a decision about each of the cards presented to you. It is alright for you to change your mind as you work through a problem, and I will not record your 'turn/don't turn' decisions until you tell me that they are your final answers.

Whilst you are reading through each problem and deciding how to solve it, please remember that I would like you to think aloud. As I've explained, you should find it quite natural to say aloud whatever happens to come into your head whilst you are working on these tasks. If you do fall silent for any length of time, however, I will gently prompt you to try and keep thinking aloud.

Participants had the chance to read through the instructions and seek clarification concerning any of the study requirements.

3.4.2 Results

Coding the inspection-time data

The same procedure for coding the data was used as in Experiments 1 to 4. Coding was halted when participants had finished making decisions about all four cards, that is, when the pointer was moved away from the final card on which a decision had been made.

Coding the verbal protocol data

Owing to technical problems during data acquisition the audio-channel on the video system failed to record the think-aloud verbalisations of participants. This problem was only identified subsequent to full data collection from all participants. As such, it was not possible to code or analyse and verbal protocol data in this experiment. Whilst unfortunate, this technical failure was not viewed as giving rise to a major problem with the experiment since its primary aim was to examine the possible existence of changes in eye-movement data that may have been engendered by the instruction to think aloud, rather than to examine changes in the think-aloud data that may have arisen from eye-movement tracking. Indeed, the eye-tracking procedure itself is highly non-invasive such that it would be most unlikely to have any reactive effect on the production of think-aloud protocols.

Statistical analysis

Descriptive analysis of the data revealed that they were positively skewed. Logarithmic transformations applying a constant of 0.8 were employed to normalise the data and to stabilise variances. For clarity of interpretation means both before transformation and converted back into their original units after transformation are reported.

<u>Card selection frequencies</u>

The same analysis to investigate the presence of a matching-bias pattern in responding was undertaken as in all previous experiments. Wilcoxon signed-ranks tests (one-tailed) revealed that more antecedent-matching cards were selected than antecedent-mismatching ones, $p = .001$, and that more consequent-matching cards were selected than consequent-mismatching ones, $p = .003$. The standard matching-bias pattern is, therefore, strongly evident in the card-selection responses associated with the present enforced-decision paradigm that also involved the deployment of eye-movement tracking and verbal-protocol methodologies.

<u>Card inspection times</u>

Selection frequencies and mean inspection times (before and after transformation) for all cards are shown in Table 3.9. The first prediction linked to the heuristic-analytic theory was tested, that is, that cards with higher selection frequencies will have higher mean inspection times. The analysis revealed a strong association between selection frequencies and inspection times, $r = .62$, $N = 16$, $p = .010$ (transformed data).

The second analysis tested Prediction 2 that derived from the heuristic-analytic theory: that for any given card, mean inspection times for individuals selecting it will be greater than for those not selecting it. Table 3.10(a) shows the mean inspection times for selected and non-selected cards for all 16 cases. The difference between mean times on selected and non-selected cards (on transformed data) was in the expected direction for 14 of the 16 cards. This was significant on the binomial test (p = .004), showing good support for Prediction 2.

TABLE 3.9

Percentage frequency of card selections and overall mean inspection times (in seconds) and log times (in seconds) for each item in Experiment 7 ($N = 31$).

		Card							
		TA		*FA*		*TC*		*FC*	
Rule		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1. If p then q	Frequency of selections	93.5		12.9		80.6		9.7	
	Natural data: time	6.73	4.33	3.88	2.42	6.53	4.99	4.57	3.08
	Transformed data: log time	0.80	0.28	0.62	0.22	0.78	0.28	0.65	0.28
	Transformed data: corrected time	5.51		3.37		5.23		3.67	
2. If p then not q	Frequency of selections	93.5		16.1		29.0		61.3	
	Natural data: time	7.88	5.54	5.74	4.87	4.57	3.10	6.55	3.98
	Transformed data: log time	0.85	0.30	0.70	0.33	0.65	0.28	0.80	0.27
	Transformed data: corrected time	6.28		4.21		3.67		5.51	
3. If not p then q	Frequency of selections	67.7		45.2		61.3		38.7	
	Natural data: time	8.06	9.54	10.35	9.91	9.74	8.18	8.30	7.50
	Transformed data: log time	0.78	0.37	0.91	0.34	0.89	0.36	0.83	0.34
	Transformed data: corrected time	5.23		7.33		6.96		5.96	
4. If not p then not q	Frequency of selections	67.7		48.4		61.3		48.4	
	Natural data: time	5.64	4.41	8.42	9.21	6.81	4.46	6.64	5.61
	Transformed data: log time	0.73	0.27	0.80	0.38	0.80	0.29	0.77	0.32
	Transformed data: corrected time	4.57		5.51		5.51		5.09	
Mean		80.6		30.7		58.1		39.5	

Note: TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

The more powerful heuristic-analytic prediction, Prediction 3, is that for each participant, mean inspection times should be longer for the cards that they have selected than for those that they have not selected. For each individual, two means were calculated from the transformed data, and a within-participants analysis of variance was performed [see Table 3.10(b)]. This confirmed Prediction 3: Inspection times were longer for the cards that were selected than those that were not, $F(1, 30) = 20.50$, $MSE = 0.35$, $p < .001$.

TABLE 3.10

Mean inspection times for selected and non-selected cards, for Experiment 7, showing (a) inspection times (in seconds) for each item, and b) inspection times (in seconds) and log times (in seconds) overall by participants ($N = 31$).

a) by items

		Selected			*Not Selected*		
			Mean			*Mean*	
Rule	*Card*	*N*	*TD*	*ND*	*N*	*TD*	*ND*
1. If p then q	TA	29	5.66	0.14	2	3.37	0.00
	FA	4	5.97	6.54	27	3.09	3.49
	TC	25	5.37	6.66	6	4.95	5.97
	FC	3	8.53	8.7	28	3.37	4.13
2. If p then not q	TA	29	6.12	7.33	2	9.67	10.10
	FA	5	9.20	10.93	26	3.67	4.74
	TC	9	4.21	5.62	22	3.47	4.13
	FC	19	6.12	7.27	12	4.57	5.42
3. If not p then q	TA	21	6.12	9.40	10	3.99	5.24
	FA	14	10.95	14.86	17	5.37	6.64
	TC	19	8.11	10.71	12	5.66	8.22
	FC	12	6.44	9.03	19	5.66	7.83
4. If not p then not q	TA	21	4.82	5.95	10	4.10	4.99
	FA	15	8.97	12.57	16	3.47	4.52
	TC	19	7.14	8.36	12	3.57	4.35
	FC	15	4.70	6.22	16	5.37	7.03

b) by participants

	Selected		*Not selected*	
	Mean	*SD*	*Mean*	*SD*
Natural data	8.48	4.26	5.39	2.60

Transformed data: log time	0.85	0.19	0.70	0.17
Transformed data: corrected time	6.28		4.21	

Note: TD = transformed data (in original units); ND = natural data; TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent.

3.4.3 Discussion

The rationale for Experiment 7 was to assess any possible reactive effects of a verbal think-aloud requirement on normal processing with abstract versions of the selection task. To explore this reactivity issue, Experiment 7 involved participants producing concurrent verbal protocols during selection-task performance whilst an eye-movement measure of their inspection-time behaviour was simultaneously recorded. It was predicted that any evidence for reactivity caused by the verbalisation requirement would reveal itself as either a switched, eradicated or largely increased inspection-time effect for selected versus rejected cards. The results revealed, however, that the basic direction and size of the inspection-time effect remained stable, with selected cards being inspected for reliably longer than rejected cards.

The existence of a largely unaltered inspection-time effect under a think-aloud requirement runs counter to any notion of reactivity on normal task processing arising from the need to produce a concurrent verbalisation. Moreover, the lack of reactivity means that the verbal protocol evidence from Experiments 5 and 6 has been substantiated: It would seem that concurrent verbal-protocol data arising in the selection task can provide an accurate index of the content and structure of ongoing thinking processes. In addition, it is important to note that the basic replication of the inspection-time effect serves, once again, to demonstrate the robustness of the effect under different instructional manipulations. The robustness of the effect seems to provide further support for Evans' (e.g., 1996, 2006) heuristic-analytic theory of the abstract selection task.

The eye-movement data also enabled an examination of a second issue, which concerned the potential role of Level 1 versus Level 2 verbalisation processes during protocol production with the selection task. In relation to this issue, it is noteworthy that inspection times for cards were generally slightly longer in duration (by a few seconds) across all cards in Experiment 7 relative to cards in Experiments 1 to 4. This intriguing observation seems to have two possible interpretations. The first is that Level 2 verbalisation is involved in the selection task, which entails a re-coding process taking place from visuo-spatial representations. This re-coding process would require a brief amount of time to achieve, thereby adding some processing time to the task. The fact that *all* of the cards are equally associated with an increase in inspection times (compared with the earlier eye-tracking experiments reported) may actually fit in with Johnson-Laird and Byrne's (2002) notion that *all* cards should be subjected to some consideration in the selection task in order to check their status in relation to the rule's truth or falsity. The trouble with taking the general increase in inspection times for all cards as support for the mental model theory, however, is that the inspection-time effect (i.e., longer consideration of selected versus rejected cards) seems itself to be *incompatible* with the model theory, since the consideration of all cards should break the observed link between longer inspection times and card selection. An alternative explanation for the general increase inspection times for all cards in Experiment 7 involves the suggestion that even a Level 1 verbalisation requirement can slow down normal thinking. Although contrary to Ericsson and Simon's (e.g., 1993) original proposals, there does appear to be a body of emerging evidence to support this proposal (Anderson, 1985; Biggs, Rosman, & Sergenian, 1993; Erber & Fiske, 1984).

Overall, then, Experiment 7 has demonstrated that engaging in the production of a concurrent verbal protocol during performance with abstract versions of the selection task does not produce any disruptive effect on the normal nature and organisation of reasoning processes—at least as detected by eye-movement analysis. This can be seen in as much as the inspection-time effect observed in Experiments 1 to 4 persists under the verbalisation instruction. This lack of any apparent disruption to primary task processing means that the verbal protocol evidence for the heuristic-analytic theory that arose from Experiments 5 and 6 seems to be substantiated. The increase in inspection times across all cards in Experiment 7 remains intriguing, although this finding is, perhaps, most compatible with the view that a think-aloud requirement, even at the status of a Level 1 externalisation of verbally-heeded information—can still slow down primary task processing. The more general issue of whether the findings from Experiments 4 to 7 can be accommodated by selection-task theories other than the heuristic-analytic account will be examined in detail in the next, and final, chapter of this thesis.

3.5 General Discussion of Experiment 4, 5, 6 and 7.

Experiments 4 to 7 were motivated by Evans' (e.g., 1996, 2006) heuristic-analytic account of matching-bias effects with abstract selection tasks. In the heuristic-analytic account, Evans claims that preconscious, heuristic processes direct attention towards cards that appear to be relevant (which end up being selected) and away from cards that appear to be irrelevant (which end up being rejected). The theory goes on to propose that where analytic reasoning is applied, it is assumed not to play a major role in determining card selections, and instead just serves to rationalise decisions already achieved on the basis of relevance. Experiments 4 to 7 were designed to investigate the small magnitude of the inspection-time effect that occurred in the eye-movement studies reported in Chapter 2, and also observed by Roberts and Newton (2001).

Experiment 4 aimed to assess whether the small inspection-time effect might be due to rule re-inspection behaviour that was clearly taking place during Experiments 1 to 3. That is, it was possible that rationalisation processes were occurring whilst participants actually re-inspected the rule that was situated at the top of each sheet during task presentation. If this was the case then an increased inspection-time effect between selected and non-selected cards might be expected if the presentation of each rule was separated from the presentation of the associated cards. Experiment 4 involved precisely this adjustment to the standard method of selection-task presentation. Yet despite the use of this rule-separation paradigm the small inspection-time effect persisted.

Once this possible methodological explanation for the small inspection-time effect had been eliminated, concurrent verbal protocols were elicited in Experiments 5 and 6 in order to explore what people think about when deliberating over cards. The verbal-protocol method is a valuable way of providing a reliable index of participant's attentional focus during task performance (Ericsson & Simon, 1993; Evans, 1989) and although it may be less sensitive than methods such as eye-movement tracking, it is able to provide an explicit trace of the content of people's thoughts whilst they engage with a task. In relation to Experiments 5 and 6 it was assumed that if it was possible to obtain a trace of the content of people's thoughts, then it should also be possible to gain a greater understanding of why the difference in inspection times between selected and non-selected cards seen in the eye-movement studies was so small in magnitude.

Experiment 5, then, used concurrent verbal protocol analysis and employed the standard selection-task paradigm requiring only active, select decisions for cards that participants thought needed to be turned over. The results provided support for the heuristic-analytic theory, showing that participants referred reliably more often to facing and hidden sides of cards that they ended up selecting compared with those that they ended up rejecting. The results provide support for previous research using verbal protocol analysis with the abstract selection task (Evans, 1995), as well as evidence for the role of secondary matching biases dominating people's references to the *hidden* sides of the cards (Wason & Evans, 1979). It is this latter finding that is of particular interest, as it suggests that people's analytic processes may be supported by the rapid, secondary cueing of matching information. This result, in fact, explains the minimal level of analytic processing on selected cards that arises in the inspection-time experiments reported in Chapter 2. Essentially, if rationalisation processes are supported by the rapid, heuristic cueing of 'relevant' values that might occur on the reverse sides of cards, then there is no reason to expect such rationalisation processes to take very long at all.

In Experiment 6, a selection-task paradigm was utilised that involved an enforced select/reject decision for all cards in order to assess what impact this decision requirement might have on the content of people's thinking. Experiment 3 in Chapter 2 has already demonstrated that matching bias and the inspection-times effects predicted by the heuristic-analytic theory are able to survive the forced-decision paradigm. However, the fact that a reduction in the magnitude of the inspection-effect was observed in Experiment 3 suggests that requiring people to attend to all cards might also have a small but detectable impact on the effect magnitudes for heuristic-analytic predictions relating to references to facing sides of cards. This was indeed seen to be the case. All facing side predictions were supported, but there was some evidence of reduction in the size of the observed effects. More importantly, however, in the case of heuristic-analytic predictions pertaining to references to *hidden* sides of cards, we expected reliable effects (as in Experiment 5), but with *no* particularly marked impact on effect magnitudes. This was because the heuristic-analytic theory would argue that people should *not* think about what values might be on the hidden sides of to-be-rejected cards (since these are judged to be irrelevant), even if the task instructions necessitate that people have to attend momentarily to the facing sides of such cards. Again, all heuristic-analytic expectations gained support from the protocol-based data obtained in Experiment 6, with reliable analytic-processing effects in evidence for selected cards versus rejected cards, and less noticeable reductions in effect magnitudes for the hidden-side predictions compared with the facing-side predictions.

Finally, Experiment 7 was carried out to assess the impact that producing verbal protocols might have on the 'normal' reasoning processes that occur on the selection task, as observed in Experiments 1 to 4. The experiment was a replication of Experiment 6, although this time participants' eye-movements were recorded alongside their verbal protocols. Combining these two methodologies produced a unique opportunity to observe whether requesting participants to generate think-aloud accounts as they tackle selection tasks modifies the way the tasks are carried out. If the eye-movements are altered then it could put into question the use of the protocol method as a reliable way in which to investigate thought processes on the selection task, as the evidence would suggest that the verbalisation requirement was having a reactive influence on normal task processing. If, however, the eye-movements are unaltered, then this is a good validation of the use of verbal protocols on the selection task and means that the results observed in Experiments 5 and 6 here, as well as elsewhere in the literature, are telling us something useful about the processes involved on the task. Suffice to say that eye-movement-based inspection-time pattern arising in the results of Experiment 7 remained *unchanged* as compared to Experiments 1 to 4, demonstrating support for the heuristic-analytic predictions and, therefore, evidence that engaging in thinking aloud whilst carrying out the selection task does not alter the natural mode of thinking that is used.

Overall, Experiments 4 to 7 have provided both eye-movement and protocol-based evidence for the role of relevance effects influencing both heuristic and analytic processing in abstract selection-task performance, as predicted by Evans' (e.g., 1996, 2006) heuristic-analytic theory. It is important, however, to consider whether other contemporary theories of the selection task are able to accommodate the present set of findings. It may well be that whilst these findings are congruent with the heuristic-analytic theory that motivated the research, they may be similarly amenable to interpretation by one or more other contemporary selection task theories. This possibility will be assessed in the next, final chapter of the thesis.

Chapter 4

General Discussion

4.1 Chapter Outline

This chapter will provide a general discussion of the experimental chapters. It will begin with a brief summary of the findings of the thesis and then go on to consider what these findings mean in relation to contemporary theories of reasoning with the selection task. This discussion will include proposals concerning how existing theories may need to be adapted in order to account for the effects reported in the present series of experiments. The chapter will then report on methodological issues arising in the course of the reported experiments that warrant further consideration, and will, in particular, concentrate on the validity of the methods that the thesis has employed, including the assumptions underpinning the methods used. Finally, the chapter will turn towards a consideration of ways in which the methodologies deployed in this thesis might profitably be applied in future research with both the Wason selection task and with other reasoning paradigms in order to gain a deeper understanding of the interplay between relevance and rationalisation processes in reasoning tasks.

4.2 Summary of Findings

The present thesis had three main aims. The first was to the attempt to improve upon previous mouse-tracking techniques that have been employed in the reasoning literature with the Wason selection task (e.g., Evans, 1996; Roberts, 1998b) by instead using eye-movement tracking as a more direct method for monitoring the moment-by-moment transitions in the locus of participants' attentional focus during reasoning. The second aim was to use the data deriving from the use the eye-tracking to establish the existence of an inspection-time effect (whereby selected cards are considered longer than rejected ones). This effect has proved to be inconsistent in mouse-tracking studies of the selection task (Evans, 1996; Roberts, 1998b; Roberts & Newton, 2001), seemingly due to task-based artefacts that can arise through the use of mouse pointing. Yet the existence of an inspection-time effect is critical for the viability of one of the dominant theories of selection-task behaviour, that is, the heuristic-analytic theory of Evans (e.g., 1984, 1989, 2006; see also Evans & Over, 1996). The third aim of the thesis was to utilise the verbal protocol method in order to investigate further the magnitude of inspection-time effects produced by the eye-tracking experiments. One particular benefit of adopting the verbal think-aloud technique in the present research programme was that it allowed the technique to be used with far more participants than have been assessed in previous selection-task experiments (e.g., Beattie & Baron, 1988; Evans, 1995).

Experiments 1 to 3 were eye-tracking experiments that set out to improve upon the previous mouse-tracking methodology that exists in the literature on the selection task. First, Evans' (1996) mouse-tracking experiment was replicated in Experiment 1 (with the slight change of using only abstract conditional materials along with the full negations paradigm) exchanging the mouse-tracking method for the eye-tracking method. Although Experiment 1 produced a highly reliable inspection-time effect, a methodological problem was identified that may have influenced the results, giving rise to artefactual support for the existence of an inspection-time effect. This artefact had arisen from the request in Experiment 1 for participants to make active select decisions only, with just a 'passive' reject response being required for non-selected cards. The concern was that this active decision requirement for selected cards might have led to inflated inspection times on such cards, thus either *creating* the inspection-time effect in the first place or *inflating* a far weaker effect.

Experiments 2 and 3 set out to remove this task-format bias, first by separating out the reasoning component of the task from the decision making section of the task (in Experiment 2), and second by equalising the influence of any potential pointing bias by utilising a forced-decision paradigm, whereby participants made a 'select/don't select' decision for all four cards. Both Experiments 2 and 3 again provided support for the inspection-time effect, indicating that it was reliable over these methodological changes aimed at removing all remaining task-format biases with the selection-task paradigm. The robustness of the inspection-time effect across all three initial experiments provided good support for the predictions of Evans' (e.g., 1996, 2006) heuristic-analytic theory, from which the inspection-time predictions were themselves derived. Thus the first two aims of the thesis were largely fulfilled through Experiments 1 to 3.

Although Experiments 1 to 3 provided reliable evidence for the inspection-time effect, a lingering issue needed further investigation. The magnitude of the effect that was observed in the first three experiments was small. Indeed, participants spent very little time inspecting cards—regardless of whether they ended up selecting them or not—and the size of the inspection-time effect was only about one-third of a second in the most unbiased experiment of all (Experiment 3). It was the need to examine the nature and cause of this small inspection-time effect that motivated the experiments that were conducted in the second half of this thesis.

Possible explanations of the small inspection-time effect could be that either: (1) the rationalisation processes that arise in the selection task are extremely quick and are driven by the rapid, heuristic cueing of information that may appear on the reverse sides of cards; or (2) that rationalisation processes on the task occur relatively slowly but at a point when participants are *re-inspecting* the rule as they were carrying out the task. The latter proposal, then, is that the inspection times that might have accumulated on to-be-selected cards were instead being distributed to other parts of the task—specifically the presented rule. Experiment 4 set out to check this methodological explanation of apparently rapid rationalisation processes by separating out the rule presentation from the card presentation in order to prevent participants thinking about card selections when they were inspecting the rule. If the small size of the inspection-time effect remained under this manipulation then this would indicate extremely quick rationalisation processes. On the other hand, if the size of the inspection-time effect was seen to increase in this rule-separation paradigm, then this would support a methodological explanation for the small inspection-time effects seen in Experiments 1 to 3. As it transpired, the small inspection-time effect persisted in Experiment 4, indicating that rationalisation processes must indeed be extremely quick, and presumably be guided by rapid heuristic processing of values on the reverse sides of cards.

Experiments 5 and 6 aimed to investigate the precise nature of the rationalisation processes observed in the previous experiments. A think-aloud reporting technique was utilised as a different process-tracing method to eye-tracking that would allow examination of the content of processes occurring on the selection task. Experiment 5 used the standard, abstract selection task requiring only active 'select' decision for to-be-selected cards. Participants were additionally asked to think aloud concurrently with their task performance. Analysis of the resulting verbal protocols revealed that participants referred reliably more often to facing and hidden sides of cards that they ended up selecting compared with those that they ended up rejecting. This finding provides converging evidence to the eye-tracking data for the role of heuristic and analytic processes in the selection task, whereby people focus their attentional processing on to-be-selected cards. The data also produced evidence for the role of secondary matching biases in the selection task (Wason and Evans, 1975), whereby people's references to the *hidden* sides of the cards are dominated by the consideration of possible 'matching' values that may appear there. This latter finding suggests that people's analytic processes may be supported by the rapid, secondary cueing of matching information on hidden card-sides.

Experiment 6 adopted the same procedure except that a forced-decision paradigm was employed in order to investigate the impact of enforced select/reject decision making for all cards on the content of people's thinking. As with the slight reduction of the inspection-time effect observed in Experiment 3, it was predicted that the same might occur within this paradigm for references to facing sides of cards. This was indeed found to be the case, that is, although all predictions were supported, there was evidence for a reduction of the size of the effect relating to references to facings sides of selected versus rejected cards. In relation to the predictions for the hidden sides of cards, however, it was expected that this effect size would remain unchanged, as according to the heuristic-analytic predictions that were being tested, people should not think about the hidden values of to-be-rejected cards. This prediction was also supported.

Finally, Experiment 7 used the eye-tracking methodology alongside a think-aloud verbalisation requirement to ensure that the verbal protocol method was not responsible for any disruption to the normal process of reasoning on the selection task. If the eye-movement-based inspection-time effect remained despite the think-aloud instruction then this gives some reassurance that no such disruption to normal processing had occurred. The persistence of the inspection time effect in Experiment 7 suggested that this was indeed the case. The slight increase in the inspection times for all cards in this experiment can be accounted for by research in the literature that suggests that Level 1 verbalisation does, in fact, slow down normal thinking a little.

One final finding to note is that in all experiments, card selection frequencies indicated not only the expected preference for TA cards over FA ones, but also a small preference for TC cards over FC ones. This finding suggests that there is some evidence for a verification bias in these experiments, as well as the standard matching bias response. Evidence of verification bias in selection tasks has been inconsistent in the literature, for example, Manktelow and Evans (1979) found no overall preference between TC and FC cards over a number of experiments, whereas the preference has been reported by other authors such as Reich and Ruth (1982) and Krauth (1982). It is not clear why verification bias is present consistently throughout the data reported in this thesis. One possibility is that verification bias is indicative of a superficial mode of responding on the task that works in conjunction with matching bias. This might arise because some participants in the present experiments were (for some reason) simply not engaging fully with the task instructions, Such a lack of engagement could promote very superficial responding – albeit responding that is sensitive to the presence of negations within rules (cf. Evans, 1995). For example, when presented with the rule '*If there is not an N on one side of the card then there is not an 8 on the other side of the card*' and the choice of cards are N, T, 1, 8, the normal matching response would be N and 8, but participants may be choosing T and 1 via a superficial response-bias mechanism that is attentive to negations (see Oaksford, 2002a, for related ideas). This is clearly an interesting possibility that would be worthy of further research.

In summary, all of the inspection-time and card-reference predictions were upheld across the seven reported experiments that involved a range of different task-format manipulations as well as the deployment of two very different process-tracing methodologies. It is now important to consider the findings of the experiments in more depth in relation to the theories of the Wason selection task discussed earlier in the thesis.

4.3 Findings in Relation to Theories of Reasoning

If we put aside for the moment the origin of the predictions of the inspection-time paradigm, the findings of the thesis were essentially: (1) a successful replication of the inspection-time effect in the selection tasks with eye-movements, in that people look longer at the cards they end up selecting than the ones they end up rejecting; (2) that the difference between the inspection times for selected versus rejected cards is small; and (3) that people refer more often to the facing and hidden sides of the cards they end up selecting than the ones they end up rejecting, and refer more to hidden matching values than hidden non-matching values. In order to explain these findings we need to apply the theories of reasoning that have been under discussion throughout the thesis.

4.3.1 The heuristic-analytic theory

The findings of the experiments reported in the thesis offer clear and strong support for the heuristic-analytic theory of reasoning as proposed by Evans (e.g., 1984, 1989, 2006; Evans & Over, 1996). Indeed, the inspection-time predictions that have been tested derived originally from this theory. The theory accounts for card selections on the Wason selection task by suggesting that people's attention is directed by preconscious, heuristic processes (i.e., the '*matching* heuristic' and the '*if* heuristic') that result in attention to cards that appear relevant and away from cards that seem irrelevant. Relevant cards get selected whilst irrelevant cards are rejected. In particular, and shown consistently in the eye-tracking experiments presented in the thesis, inspection times are longer for the selected matching *and* mismatching cards over rejected ones. This is because relevance effects extend beyond just the matching cards in determining card selections, since the *if*-heuristic encourages selection of the TA card across all four rules types. This means that the matching TA cards will be selected on two of the rules and the mismatching TA cards will be selected on the other two rules. This finding is consistent with the heuristic-analytic theory as card selections are deriving from judgements of relevance.

Conscious and rational processes that occur on the task only serve to rationalise card decisions that have already been made on the basis of relevance. As stated above, this proposal is clear in the inspection-time effect that is established across all the eye-tracking experiments. Indeed, it is the rationalisation process that is responsible for the inspection-time effect, as rationalisation is deployed to justify only select decisions and not reject decisions.

The verbal-protocols experiments give further support for the role of analytic rationalisation processes on the selection task, as they show that people also make more spoken references to the facing and hidden sides of the cards they end up selecting. The support comes especially from the finding regarding references to the hidden sides of the cards. In particular we can see that although the hidden sides of the cards are considered, this does not have any effect on card selections (if it did we would expect the link between references to the hidden sides of the cards and the card selections to be severed; that is, people might spend time referring to a card but end up *rejecting* it). Indeed, instead of playing a role in determining card choices, thinking about the other side of the card instead seems to confirm card selections—selections already made on the basis of relevance.

The verbal protocol experiments provide even more evidence for the heuristic-analytic account if we consider what it is people are actually thinking about when they consider the reverse sides of the cards. They are clearly not thinking about the mismatching values on hidden sides, and, according to the heuristic-analytic account, this is because people do not see the mismatching cards as relevant to the decision-making process. What we have seen in the protocols in relation to references to the hidden sides of the cards is that such references are dominated by the mention of matching possibilities. This secondary matching bias that appears to be taking place serves to cue the analytic processes into a justification of selecting those particular cards (see also Wason & Evans, 1975). The secondary matching bias is also an explanation for the small inspection-time effect that persists throughout the experiments. We have explained this small effect as being due to a rapid rationalisation process that would, indeed, be very quick if people's justifications are determined by the heuristic cueing of relevant (i.e., matching) values that could appear on the reverse sides of cards.

As can be seen here, then, as well as in earlier discussions, the heuristic-analytic account of selection-task performance has been supported by the use of eye-tracking and verbal-protocol methodologies, and, as such, provides a good account of reasoning processes on the task. Indeed, Evans' (e.g., 2003, 2006) more recent proposals of his hypothetical thinking theory and extension of the heuristic-analytic theory is now nearing an almost step-by-step model of the interplay between heuristic and analytic processes in reasoning. This is particularly the case with Evans' (2006) recent introduction of the three principles of hypothetical thinking into his heuristic-analytic theorising, that is, the singularity principle (people construct only one mental model at a time in which to represent a hypothetical situation), the relevance principle (people consider the model which is the most relevant in the context) and the satisficing principle (people evaluate models according to their current goals and accept models that appear to be satisfactory). The satisficing principle provides an especially good account of why the analytic system typically fails to override heuristically-cued choices in the selection task: Most people's analytic systems will simply satisfice (accepting heuristically-cued choices) whenever *either* a verification or a falsification justification for selecting a card can be found (depending on the matching value on the hidden side of the card). In practice, this means that heuristically-cued choices on the selection task will almost invariably be accepted.

Essentially, then, Evans (2006) is arguing that heuristic processes bias and shape analytic reasoning by cueing default mental models that lead to default decisions. In these more recent proposals, Evans suggests that the intervention of analytic processing may or may not occur to inhibit this default responding by revision or replacement of the models. The sort of situation that analytic intervention will occur in is where the reasoner has very high cognitive ability (i.e., high working memory capacity), where people are instructed to engage in logical reasoning, or where there is more time for people to engage in reflective thinking.

4.3.2 The information gain theory

The second theory to consider is Oaksford and Chater's (e.g., 1994, 1996, 2003) information gain account. As noted in Chapter 1, the information gain theory has a compelling track record in terms of its capacity to explain many aspects of selection-task performance (including the influence of probabilistic manipulations) across a variety of task variants. According to the theory, information gain provides a formal measure of 'relevance' (see Oaksford & Chater, 1995), and, therefore, information gain appears to predict the same basic pattern of matching-card selections as envisaged by the heuristic-analytic theory.

In relation to the findings in this thesis, the theory has mixed applicability. With regards to both the inspection-time paradigm and the verbal-protocol paradigm, the information gain theory appears at first sight to predict identical inspection times or facing side references across all cards since reasoners need to undertake expected information gain computations on each card to determine its potential support for the rule. This clearly is not supported by the findings of this thesis. On closer analysis, however, information gain theory may well be able to provide an alternative account of the findings that is distinct from Evans' (e.g., 1984) emphasis on the linguistic basis of matching effects (I am grateful to Oaksford, personal communication, for alerting me to this). So, for example, relevance assessments determined on-line by participants via information-gain calculations *could* lead to more references to matching versus mismatching values on facing sides, essentially because people will end up showing a greater level of interest in the relevant cards (i.e., the 'rare' items) than the irrelevant ones—which would also contribute to an inspection-time effect. In a similar way, information gain theory would also predict that in justifying their card selections participants would show *secondary matching* because for all rules they are searching for the *rare* cases (which are always the matching antecedent and matching consequent combination).

Overall, then, the information gain model may well be able to capture the relevance effects that we have demonstrated in relation to references to both facing and hidden card sides, as well as the basic inspection-time effect established through eye-movement tracking. This theory, moreover, would describe secondary matching effects arising in references to hidden sides of cards as analytic response (Oaksford, personal communication). This account also ties in selection-task behaviour to rational explanations of biases in judgements relating to 2 x 2 contingency tables (e.g., Anderson & Sheu, 1995; Over & Green, 2001).

One potential weakness with the information gain account as it is currently formulated, however, is that it does not provide a full-blown *algorithmic* level theory specifying the specific nature, organisation and time-course of the processing steps underpinning card selections (i.e., it is formulated at the *computational* level of what is being computed). Indeed, to derive an account of both our inspection-time data and the verbal-protocol data we have had to go quite some way beyond the assumptions of the theory as currently explicated in published research. As Oaksford (personal communication) has pointed out, however, most current models of the selection task (and not just the information gain theory) are also highly underspecified in terms of the detailed operation sequences underlying reasoning, with theories simply tending to refer to a loose binary processing distinction (e.g., heuristic then analytic; initial representation then fleshing out). Indeed, the information gain model seems to be at least as capable as some other theories of affording an understanding of algorithmic level issues in the abstract selection task—as has been outlined above. Nonetheless, it would be appealing to see the information gain theory developed further at an algorithmic level; such developments are apparently underway (e.g., Oaksford, 2002b). Their fruition is certainly something to be looked forward to.

4.3.3 The mental model theory

The final theory to consider is the most recent mental models account of the selection task (Johnson-Laird & Byrne, 2002). In terms of the inspection time effects, it would seem that mental model theory would predict equivalent inspection times for all cards. This is because reasoners attempt to integrate *each* card with their model of the conditional to determine its impact on the rule's truth or falsity. The observed inspection-time difference between selected and rejected cards does not, therefore, emerge directly from mental model theory as currently stated. One way in which the theory could be reconciled with the inspection-time effect is to propose that cards that can be integrated with existing models are subjected to increased processing—perhaps whilst the logical consequences of such integration are determined—relative to cards that *cannot* be integrated with existing models.

A remaining difficulty for mental model theory, however, is to account for verbal-protocol evidence that people think primarily about matching values on hidden card sides rather than reflecting on potential falsifying and verifying values. This apparent asymmetry in what values people consider as being present on the reverse sides of cards does not readily seem to emerge from the mental models assumption that people assess cards in terms of how their hidden values might impact on the truth or falsity of the presented rule. It may well be that mental models theorists could develop a viable account for such secondary matching effects, but it remains the case that these effects were directly predicted by the heuristic-analytic theory.

It is also finally worth noting here that the secondary matching evidence uncovered with the verbal protocols are also a challenge for Feeney and Handley's (e.g., 2000) claims to have detected a deductive component in abstract variants of the selection task—a conclusion that they base on their finding that participants consider the hidden sides of presented cards. However, if when considering such hidden values most people are simply engaging in a secondary matching process, then this would seem to be evidence against deduction being a key component of reasoning in the selection task. The limited support for mental models predictions deriving from evidence for secondary-matching effects also calls into question Evans' past proposals (e.g., Evans & Over, 1996, p. 136) that mental modelling may supply the analytic component to the heuristic-analytic theory, which has always been less well specified than the heuristic component in this account. On balance, it would seem that either Evans' heuristic-analytic account (minus a mental-models analytic stage) or Oaksford and Chater's information gain theory are most readily able to explain the full breadth of protocol-based evidence that we have uncovered for relevance effects and rationalisation processes in the selection task.

In summary, Evans' (2006) recent extension of the heuristic-analytic theory, which now links to three principles of hypothetical thinking, arguably takes the heuristic-analytic account one step closer toward a detailed process-model of the indicative selection task than either mental models theory or information gain theory. Moreover, the processing steps that the heuristic-analytic theory specifies seem currently to provide the most convincing account of the small but reliable inspection-time effect demonstrated in the eye-tracking experiments and the secondary matching bias effects uncovered in the verbal protocol experiments reported in this thesis. If the information gain theory evolves to produce an algorithmic level theory then it may well end up having the edge on the heuristic-analytic account, because of its impressive ability to explain a wide range of probabilistic influences on card-selection patterns.

4.4 Methodological Issues

Now that the findings of the thesis have been considered in relation to theories of reasoning, methodological weaknesses or strengths with the reported experiments can be examined. All of the experiments in the thesis involve techniques and paradigms that have considerable potential to be reactive to methodological issues that may induce confounds and biases in the resulting data. However, the research undertaken has been very alert to such biasing possibilities at every step of the way, such that it is quite hard to identify any remaining task-based or technique-based factors that might be having any biasing effect on the results as they stand.

Two main methodological issues were dealt with over the course of the experiments. Within the first eye-tracking experiment (Experiment 1), a potentially major bias was identified, which was the requirement in the standard abstract selection task for only active decisions to be registered for card selections. This factor has a potentially large implication for the inspection-time effect: At best any cards that are selected might cause an inflated inspection-time effect; at worst any selected cards could have caused the inspection-time effect to occur in the first place. Experiments 2 and 3 were able to combat this problem, first by removing the pointing time for the selection decision (in Experiment 2), and second, by equalising the active-decision time across all four cards (in Experiment 3). With these safeguards in place a reliable inspection-time effect was still observed. Experiment 4 aimed to ensure that the small magnitude of the inspection-time effect was not due to participants thinking about cards for selection whilst engaging in re-reading of the rule. This experiment removed the rule presentation from the card presentation in order to overcome this potential problem. It turned out that the small inspection-time effect remained, despite these steps being taken, and the thesis went on to explore the processes involved in selection task performance using verbal protocols. However, this final methodological check in Experiment 4 ensured that all apparent methodological explanations of the inspection-time effect had been tested and rejected.

Beyond these paradigm-specific methodological concerns, however, there are some more wide-reaching methodological issues that remain which relate to the basic assumptions underpinning the use of eye-movement tracking and think-aloud methods in reasoning research. The justification for the use of these methods was forwarded earlier in the thesis, but will be discussed again here in the light of the findings as a whole. The eye-tracking methodology was utilised as it was believed that it provided a more direct measure of on-line attentional processing than the mouse-tracking method upon which the thesis was originally based. The eye-tracking methodology has been used extensively in a number of different information processing tasks in psychology in order to explore the underlying cognitive processes that are occurring, with the basic idea being that eye-movements reflect these moment-to-moment cognitive processes. Although we often look at something in our environment when we attend to it, it is of course possible to attend to an item without looking at it, and research has shown that the relationship between gaze direction and attention will depend on the nature of the task and its demands on the attentional 'spotlight' (Posner, 1980). If the stimuli are complex it will be more effective to move the eyes instead of one's attention (He & Kowler, 1992), and although the locus of attention and eye location can be separated in simple discrimination tasks (see Posner, 1980), it is generally agreed that in more complex information processing task such as reading, the link between the two is actually quite tight.

The eye-movement methodology has been used extensively in reading research where researchers have been interested in the nature of eye-movements such as saccades (continual ballistic movements made by the eyes), fixations (moments when the eyes remain relatively still for about 200-300 ms) and pursuit movements (when one's eyes follow a moving target); see Rayner (1998) for a review. This research deals mainly with what information (if any) is being processed across these different categories of eye movement. For the purpose of the present thesis, however, the interest in relation to eye-movements was exclusively concerned with gaze duration, that is, *where* a person spends their time looking whilst tackling a particular task. Just and Carpenter (1976) found that on tasks such as mental rotation, sentence verification and quantitative comparison, the time that people spent gazing at a figure reflected the time it took to encode and operate upon that figure. They suggested two reasons why gaze duration may continue. They propose that either fixation continues on a figure despite the relevant information having been already encoded, because the processor is busy, and so there is no need to move the eyes elsewhere, or there is an active instruction to the eye to remain where it is because saccadic movement initiates new encoding which the processor cannot deal with until it has finished dealing with previous information that has been encoded. Either way, Just and Carpenter argue that duration of gaze provides a very good measure of time spent processing stimuli.

Moving on to a consideration of the assumptions associated with the verbal-protocol method, we note that the verbalisation technique adopted for this thesis required participants to give *think aloud* accounts of their reasoning as they worked through the selection tasks presented to them. According to Ericsson and Simon (1993), this particular method should appease the critics of the use of verbal protocol methods in cognitive research who claim that verbally-based thoughts do not necessarily correlate with observable behaviour. Ericsson and Simon suggest that it is only the use of *concurrent* think-aloud reports that allows researchers to gain access to the sequential states of thought that each contain an end-product of cognitive processing. As these products are stable it is possible to produce a verbalisation of them, and these verbalisations are largely unchanged by this verbal-production process itself. There is a good array of evidence to support this idea (see Ericsson & Simon, 1993, for a review).

In summary, both eye-tracking and think-aloud techniques are well-respected methods for the collection of process-tracing data, and there is a large body of research that gives us confidence in the use of these methods within the context of this thesis. These methodologies, when used in combination, have provided a unique opportunity for a detailed investigation into relevance and rationalisation processes in the Wason selection task. Moreover, the datasets deriving from the deployment of these distinct methods appear to converge to provide a coherent set of findings that allow for meaningful theoretical interpretations of results in terms of current reasoning theories.

4.5 Future Directions

The findings of the thesis have been discussed, and it is now important to turn to the future directions of this work. It is particularly interesting to consider possible applications of the methods used here in order to understand more fully the role of relevance and rationalisation in reasoning. To this end, we note that eye-tracking methodology and verbal-protocol techniques could be applied to both different versions of the selection task as well as to a variety of other different reasoning tasks beyond the selection task. We will now consider some of these ideas in more detail.

The first obvious application of the methodologies used here would be to thematic variants of the selection task. When Evans (1996) ran his pioneering mouse-tracking experiments he uncovered an inspection-time effect for a series of thematic (including deontic) selection tasks, indicating that relevance plays a strong role in reasoning that can cut across both abstract and thematic problem contents. On thematic versions of the selection task matching bias is usually suppressed and these versions can give rise to the facilitation of the logically correct card-selection responses of *p* and *not-q*. The heuristic-analytic account would argue that on these tasks, facilitation occurs due to the presence of alternative cues to relevance, such as pragmatic cues instead of the linguistic cues that operate on the indicative version. Such pragmatic cues seem to operate providing that there is: (1) familiarity with the rule (or at least that the rationale of the rule for guiding behaviour is clear); and (2) some minimal context to the rule (removal of minimal content has been shown to have detrimental effects on facilitation; see Evans & Pollard, 1987).

On the basis of the findings of this thesis and those reported by Evans (1996), it is predicted that the application of the eye-movement and verbal protocol methodologies to thematic versions of the selection task would produce very similar findings relating to the processing times for selected versus rejected cards as well as to the frequency of references to the facing and hidden sides of such cards. Such a generalisation of the effects reported in the thesis would be important as it would consolidate the findings of this thesis and give further support to the heuristic-analytic account. In addition, this generalisation might also help to arbitrate between the heuristic-analytic account and other reasoning theories such as the information gain account.

A further use for the eye-tracking methodology with the selection-task paradigm would be to deploy it in studies using a rapid-response version of the selection task, as developed by Roberts and Newton (2001). In their Experiment 2, Roberts and Newton presented participants with different versions of the selection task (both abstract and thematic). The free time group were presented with a blanked-out preamble and rule as well as blanked-out cards. Viewing each part of the task involved holding the mouse pointer over one of these areas, and as this was done the card became visible and card inspection times were calculated. Participants had to make a decision about all four cards but could view each of the cards as many times as they liked. The rapid response group had a similar set up, being only able to view one card at a time, or the preamble and the rule. However, participants had *only one second* to view each card and a further second to make their decision about that card. The prediction was that if analytic processing was having no effect on card selections—which were being driven purely by heuristic processes—then there should be no difference in the selections between the free time and the rapid response groups. The findings were broadly in line with this prediction, although the card-selection patterns did actually show a slight increase in consequent-matching decisions for the rapid response task. The latter finding suggests that analytic processes for some participants in the free time group may well have been functional in overturning heuristic cueing of matching consequent cards. For the purpose of the present discussion, however, it is notable that Roberts and Newton did not collect any response times for the rapid response task (presumably because the time-limitations associated with the task were so restrictive). But if a more standard presentation format was adopted for the selection task then eye-movement tracking could reveal interesting aspects of on-line processing even under rapid response instructions—where, say, participants had just 10 s to complete the whole task. Based on Roberts and Newton's results, it would be predicted that a time restriction would allow no opportunity for rationalisation to occur, which would result in no inspection-time effect, but the same standard pattern of matching card selections should persist because of the dominance of heuristic processes.

An additional, interesting application of the combined deployment of eye-tracking and verbal-protocol methodologies would be in relation to the experimental manipulations presented by Feeney and Handley (2000). They successfully obtained *q* card suppression on a version of the selection task that involved the presentation of two conditional rules that contained alternative antecedents to one another. They claim that when using this rule manipulation people are able to recognise that on the back of the *q* card there may be a *p* case or a *not-p* case, and this then leads them to infer that the *q* card actually has no bearing on the truth or falsity of the rule. Feeney and Handley suggest that this effect could only arise if people are able to consider explicitly the reverse sides of the cards. If this were the case, then using the inspection-time paradigm with this manipulation would lead to the prediction of an increase in the time that participants spend looking at the *q* card—but with the subsequent rejection of this card. This would break the link between increased card inspection and card selection such that no inspection-time effect would be uncovered. Essentially, then, it may be that the very nature of this version of the selection task—that is, the extra antecedent in the second conditional rule—is causing other cards to appear relevant, and so the pattern of selections is altered. The claim by Feeney and Handley that people are considering what is on the back of the *q* card, could clearly be investigated further with the application of the verbal-protocol method. The application of both eye-tracking and verbal protocol analysis to the Feeney and Handley paradigm would allow some of these issues regarding the presence of deductive reasoning versus the dominance of rationalisation and relevance to be clarified more fully.

Although there are many other potential applications of the eye-tracking and verbal-protocol methodologies to the selection task, the last one that will be considered here is in relation to probabilistic manipulations on the selection task, in particular to deontic versions of the task (e.g., Manktelow & Over, 1991). One such study by Manktelow, Sutherland, and Over (1995), used an enlarged-array selection task to investigate the role of probabilistic factors in reasoning with conditional obligations. This study is of interest because it added probabilistic information (information that causes items to appear more relevant) to both antecedent and consequent items in the Cheng and Holyoak (1985) immigration task, where the rule specifies that '*If a passenger's form says ENTERING on one side then the other side must include cholera amongst the list of diseases*' (see Chapter 1). Manktelow et al.'s manipulation increased the number of cards used in the task from four, in the original version, to 20. Six of these were ENTERING (*p*) cards, four said TRANSIT (*not-p*), five included cholera among a list of diseases (*q*) and five did not include cholera among a list of diseases (*not-q*). On the *p* and *not-p* cards, probabilistic information about the passenger's country of origin was inserted, that is, on half the cards a tropical country was inserted (e.g., Thailand, where cholera is present), and on the other half a European country was inserted (e.g., Denmark, where cholera is not present). Participants were, therefore, presented with extra relevant information about those passengers that present a higher risk.

Manktelow et al. (1995) predicted that the *p* and *not-q* cards would be the most frequently selected, but that participants would select more of the cards that indicated countries of high infection probability and the specified disease, compared with cards for travellers from areas with a lower probability of infection. The results showed that the *p* cards were indeed affected by probability information, in that there was a suppression of *p* cards if they had a European country on them. Manktelow et al. propose that card selections on this task are being mediated by differing subjective expectations of low-probability and high-probability items, in particular, low-probability items are deemed to be of less value than high-probability ones. Within the eye-movement based inspection-time paradigm this finding could be explored further. It would be interesting, for example, to determine whether people spend time looking equally across all *p* cards, or whether they spend more time looking at the *p* cards that hold more 'utility' or appear more 'relevant'. Indeed, it would be possible to undertake a detailed inspection-time analysis that was specifically focused on just the *p* cases to determine if there were longer looking times at selected *p* cards than rejected *p* cards.

Finally, we note that relevance effects are reported widely elsewhere in the literature on conditional reasoning, particularly on truth-table tasks. One type of relevance effect can be seen in studies of both truth-table evaluation tasks (e.g., Johnson-Laird & Tagart, 1969) and truth-table construction tasks (e.g., Evans, 1972) that show that participants judge *false antecedent* cards to be 'irrelevant'. A second type of relevance effect in truth-table tasks is *matching bias*; indeed it was the truth-table construction task where matching bias was first discovered (Evans, 1972), as demonstrated by the finding that people are more likely to construct a case that matches either or both components of the conditional. These relevance effects, as we know, are also demonstrated in the selection task.

Essentially, then, the truth-table task and the selection task both provide evidence for two different sources of relevance, that is: (1) cases appear more relevant when the antecedent condition is fulfilled, and (2) cases appear more relevant when their features match those named in the conditional (Evans & Over, 1996). The interesting difference between matching bias on the truth-table task and matching bias on the selection task is that when matching bias is eliminated on both tasks (e.g., Evans, 1983; Evans, Clibbens, & Rood, 1996), there is an increase in logical responding on the truth-table task but *not* on the selection task. Evans proposes that this is because the truth-table task elicits analytical as well as heuristic processes, whereas, as we have seen, analytic processes on the selection task do nothing other than to rationalise card selections. Such differences between truth-table tasks and selection tasks could readily be investigated further with the use of both eye-tracking and verbal-protocol methods.

In summary, there are a number of predictions arising from the heuristic-analytic theory of thematic and probabilistic selection tasks as well as truth-table tasks that could profitably be examined using eye-movement inspection-time measures as well as verbal-protocol assessments of the content of people's thought processes. All of these predictions are easily testable. It is possible that the findings from such experiments would, in particular, add to the growing experimental evidence for the presence of relevance and rationalisation processes in the selection task. Such evidence could, itself, give increased support for the heuristic-analytic theory of reasoning, and thereby make an important, further contribution to current debates concerning the nature of human reasoning process.

4.6 Summary statement

The experiments presented in this thesis have demonstrated that valuable data can be collected from a combination of different process-tracing methods, such as eye-movement tracking and think-aloud protocol analysis. Such data have allowed for an investigation of the abstract Wason selection task that goes well beyond simple card-selection patterns, instead enabling the processes behind these selections to be examined in detail. The data have also helped to inform an understanding of different theories of reasoning, and suggest that the heuristic-analytic account (e.g., Evans, 1996, 2006) may currently be the leading theory of selection-task performance, as it is the only theory that is sufficiently well-specified to be able to account readily for the full set of process-oriented findings uncovered in the present research programme.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14*, 471-485.

Anderson, J. R. & Sheu C.-F. (1995). Causal inferences as perceptual judgements. *Memory and Cognition, 23*, 510-524.

Anderson, M. (1985). Some evidence of the effect of verbalization on process: A methodological note. *Journal of Accounting Research, 23*, 843-852.

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology*, *53*, 77-86.

Baron, J. (2000). *Thinking and deciding*. Cambridge: Cambridge University Press.

Beattie, J. & Baron, J. (1988). Confirmation and matching biases in hypothesis testing. *Quarterly Journal of Experimental Psychology, 40A*, 269-297.

Biggs, S. F., Rosman, A. J., & Sergenian, G. K. (1993). Methodological issues in judgement and decision-making research: Concurrent verbal protocol validity. *Journal of Behavioural Decision-Making, 6*, 187-206.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT Press.

Cheng, P. W. & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

Cheng, P. W. & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition, 33,* 285-313.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, R. M. (1986). Pragmatic versus syntactic aproaches to training deductive reasoning. *Cognitive Psychology, 18,* 293-328.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317-370.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the selection task. *Cognition, 31,* 187-276.

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barlow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind.* Oxford: Oxford University Press.

Craik, K. (1943). *The nature of explanation.* Cambridge: Cambridge University Press.

Dawkins, R. (1976). *The selfish gene.* Oxford: Oxford University Press.

Dawkins, R. (1996). *Climbing mount improbable.* New York: Norton and Co. Inc.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36,* 1827-1837.

Ellis, M. C. (1991). *Linguistic and semantic factors in conditional reasoning.* Unpublished PhD thesis, University of Plymouth.

Erber, R. & Fiske, S. T. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology, 47,* 709-726.

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Ericsson, K. A. & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Cambridge, MA: MIT Press.

Espino, O., Santamaria, C., Meseguer, E., & Carreiras, M. (2005). Early and late processes in syllogistic reasoning: Evidence from eye-movements. *Cognition, 98,* 1-9.

Evans, J. St. B. T. (1972). Interpretation and matching bias in a reasoning task. *British Journal of Psychology, 24*, 193-199.

Evans, J. St. B. T. (1975). On interpreting reasoning data: A reply to Van Duyne. *Cognition, 3*, 387-390.

Evans, J. St. B. T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29,* 297-306.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning.* London: Routledge.

Evans, J. St. B. T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology, 35A*, 635-644.

Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 451-468.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T. (1991). Theories of human reasoning: The fragmented state of the art. *Theory and Psychology, 1*, 83-105.

Evans, J. St. B. T (1995). Relevance and reasoning. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 147-172). Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology, 87*, 223-240.

Evans, J. St. B. T. (1998a). Inspection times, relevance and reasoning: A reply to Roberts. *Quarterly Journal of Experimental Psychology, 51A*, 811-814.

Evans, J. St. B. T. (1998b). Matching bias in the selection task: Do we understand it after 25 years? *Thinking and Reasoning, 4*, 45-82.

Evans, J. St. B. T. (2002). Matching bias and set sizes: A discussion of Yama (2001). *Thinking and Reasoning, 8*, 153-163.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454-459.

Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review, 13*, 378-395.

Evans, J. St. B. T., Ball, L.J., & Brooks, P. G. (1987). Attentional bias and decision order in a reasoning task. *British Journal of Psychology, 78*, 385-394.

Evans, J. St. B. T., Clibbens, J. and Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language, 35*, 392-409.

Evans, J. St. B. T., & Handley, S. J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology, 52A,* 739-769.

Evans, J. St. B. T., Handley, S. J. & Harper, C. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology. 54A,* 935-958.

Evans, J. St. B. T., Legrenzi, P. and Girotto, V. (1999). The influence of linguistic form on reasoning: The case of matching bias. *Quarterly Journal of Experimental Psychology, 52A,* 185-216.

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64,* 3, 391-397.

Evans, J. St. B. T., & Manktelow, K. I. (1979). Facilitation of reasoning by realism: effect or non-effect? *British Journal of Psychology, 70,* 477-488.

Evans, J. St. B. T., & Newstead, S. E. (1980). A study of disjunctive reasoning. *Psychological Research, 41,* 373-388.

Evans, J. St. B. T., Newstead, S. E. & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T, & Over, D. E. (1996). *Rationality and reasoning.* Hove: Psychology Press.

Evans, J. St. B. T, & Over, D. E. (1997). Rationality in reasoning: The case of deductive competence. *Current Psychology of Cognition, 16,* 3-38.

Evans, J. St. B. T, & Over, D. E. (2004). *If.* Oxford: Oxford University Press.

Evans, J. St .B. T., Over, D. E. & Handley, S. J. (2003). A theory of hypothetical thinking. In D. Hardman & L. Maachi (Eds). *The psychology of reasoning and decision making* (pp. 3-21). Chichester: Wiley.

Feeney, A. & Handley, S. H. (2000). The suppression of q card selections: Evidence for deductive inference in Wason's selection task. *Quarterly Journal of Experimental Psychology, 53A,* 1224-1243.

Fisher, M. H. (1999). An investigation of attention allocation during sequential eye movement tasks. *Quarterly Journal of Experimental Psychology, 52A,* 649-677.

Gigerenzer, G. & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127-171.

Girotto, V., Blaye, A., & Farioli, F. (1989). A reason to reason: Pragmatic basis of children's search for counterexamples. *European Bulletin of Cognitive Psychology, 9,* 297-321.

Goodwin, R. Q. & Wason, P. C. (1972). Degrees of insight. *British Journal of Psychology, 63,* 205-212.

Grant, E. R. & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14,* 462-466.

Green, C. (1992). *Mental models, counterexamples and the selection task.* Unpublished manuscript, Department of Psychology, University College London.

Green, C. & Larkin, R. (1995). The locus of facilitation in the abstract selection task. *Thinking and Reasoning, 1,* 183-199.

Green, C. & Over, D. E. (1997). Causal inference, contingency tables and the selection task. *Current Psychology of Cognition, 16,* 459-487.

Green, C., Over, D. E., & Pyne, R. (1997). Probability and choice in the selection task. *Thinking and Reasoning, 3,* 209-236.

Griggs, R.A. and Cox, J.R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407-420.

Griggs, R.A. and Cox, J.R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology, 35A,* 519-533.

Handley, S. J., & Feeney, A., & Harper, C. (2002). Alternative antecedents, probabilities and the suppression of fallacies in the Wason selection task. *Quarterly Journal of Experimental Psychology, 55A,* 799-813.

He, P., & Kowler, E. (1992). The role of saccades in the perception of texture patterns. *Vision Research, 32,* 2151-2163.

Hiraishi, K. & Hasegawa, T. (2001). Sharing-rule and detection of free riders in co-operative groups: Evolutionarily important reasoning in the Wason selection task. *Thinking and Reasoning, 7,* 255-294.

Hodges, W. (1993). The logical content of theories of deduction. *Behavioural and Brain Sciences, 16,* 353.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics, 57,* 787-795.

Holland, J., Holyoak, K., Nisbett, J., & Thagard, P. (1986). *Induction: Processes of inference, learning and discovery.* Cambridge, MA: MIT Press.

Holyoak, K. & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Jackson, S. L. & Griggs, R. A. (1990). The elusive pragmatic reasoning schema effect. *Quarterly Journal of Experimental Psychology, 42A,* 353-374.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (1985). Deductive reasoning ability. In R.J. Sternberg (Ed.), *Human abilities: An information processing approach.* New York: W.H. Freeman.

Johnson-Laird, P. N. (1995). Inference and mental models. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 115-146). Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Science, 5,* 434-442.

Johnson-Laird, P. N. & Byrne, R.M.J. (1991). *Deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P. N. & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inferences. *Psychological Review, 109,* 646-678.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V. & Legrenzi, M. S. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review, 106(1),* 62-88.

Johnson-Laird, P. N., Legrenzi, P. & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395-400.

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology, 2,* 367-373.

Johnson-Laird, P. N. & Wason, P. C. (1970). Insight into a logical relation. *Quarterly Journal of Experimental Psychology, 22,* 49-61.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8,* 441-480.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition, 51,* 1-28.

Klein, R. M., Kingstone, A., & Pontefract, A. (1992). Orienting of visual attention. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 46-65). New York: Springer-Verlag.

Knoblich, G., & Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory and Cognition, 29,* 1000-1009.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35,* 1897-1916.

Krauth, J. (1982). Formulation and experimental verification of models in prepositional reasoning. *Quarterly Journal of Experimental Psychology, 34A,* 285-298.

Krauth, J. & Berchtold-Neumann, M. (1988). A model for disjunctive reasoning. *Zeitschrift für Psychologie, 196,* 361-370.

Kroger, J. K., Cheng, P. W. & Holyoak, K. J. (1993). Evoking the permission schema: The impact of explicit negation and a violation checking context. *Quarterly Journal of Experimental Psychology, 46A,* 615-635.

Liberman, N. & Klar, Y. (1996). Hypothesis testing in Wason's selection task: Social exchange, cheating detection or task understanding. *Cognition, 58,* 127-56.

Liversedge, S. P., Paterson, K. B., & Pickering, M. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception*. Oxford: Elsevier Science.

Love, R. & Kessler, C. (1995). Focussing in Wason's selection task: Content and instruction effects. *Thinking and Reasoning, 1*, 153-82.

Manktelow, K. I. (1999). *Reasoning and thinking*. Hove, UK: Psychology Press, Ltd.

Manktelow, K. I. & Evans, J. St. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology, 70*, 477-488.

Manktelow, K. I. & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39*, 85-105.

Medvedev, G. (1991). *The truth about Chernobyl*. London: Tauris.

McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrable, R. P. (2001). Do conditional hypotheses target rare events? *Organizational Behavior and Human Decision Processes, 85*, 291-309.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review, 7*, 360-366.

Miles, J. N. V. & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London: Sage Publishers Ltd.

Morley, N. J., Evans, J. St. B. T. & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 57A*, 666-692.

Mugglestone, M. (1999). *EMAT: Eye-movement analysis tool*. Bedford, MA: Applied Science Laboratories.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-295.

Nodine, C. F., Kundel, H. L., Toto, L. C., & Krupinski, E. A. (1992). Recording and analyzing eye-position data using a microcomputer workstation. *Behavior Research Methods, Instruments, and Computers 24,* 475-485.

Oakhill, J., & Garnham, A. (1994). *Thinking and reasoning.* Oxford: Blackwell Publishers Inc.

Oaksford, M. (2002a). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking and Reasoning, 8*, 135-151.

Oaksford, M. (2002b). Predicting the results of reasoning experiments: Reply to Feeney and Handley (2000). *Quarterly Journal of Experimental Psychology, 55A,* 793-798.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608-631.

Oaksford, M. & Chater, N. (1995). Information gain explains relevance which explains the selection task. *Cognition, 57,* 97-108.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381-391.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*, 349-357.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review and re-evaluation. *Psychonomic Bulletin and Review, 10,* 289-318.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5*, 193-244.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory and Cognition, 23,* 441-458.

Oaksford, M. and Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology, 18,* 835-854.

Oaksford, M. and Wakefield, M. (2003). Data selection and natural sampling: Probabilities do matter. *Memory and Cognition, 18,* 835-854.

Over, D. E. & Green, D. W. (2001). Contingency, causation and adaptive inference. *Psychological Review, 108,* 682-684.

Pollard, P. & Evans, J. St. B. T. (1987). On the relationship between content and context effects in reasoning. *American Journal of Psychology, 100,* 41-60.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32A,* 3-25.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Reber, A. S. (1993). *Implicit learning and tacit knowledge.* Oxford: Oxford University Press.

Reich, S. S. & Ruth, P. (1982). Wason's selection task: Verification, falsification and matching. *British Journal of Psychology, 73,* 395-405.

Reuter-Lorenz, P. A., & Frendrich, R. (1992). Oculomotor readiness and covert orienting: Differences between central and peripheral cues. *Perception and Psychophysics, 52,* 336-344.

Roberts, M.J. (1998a). How should relevance be defined? What does inspection time measure? A reply to Evans. *Quarterly Journal of Experimental Psychology, 51A,* 815-817.

Roberts, M. J. (1998b). Inspection times and the selection task: Are they relevant? *Quarterly Journal of Experimental Psychology, 51A,* 781-810.

Roberts, M. J. (2002). The elusive matching bias effect in the disjunctive selection task. *Experimental psychology, 49,* 89-97.

Roberts, M. J. & Newton, E. J. (2001). Inspection times, the change task, and the rapid-response selection task. *Quarterly Journal of Experimental Psychology, 54A,* 1031-1048.

Roth, E. M. (1979). Facilitating insight into a reasoning task. *British Journal of Psychology, 70,* 265-272.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle, G. (1999). Is there any need for eye-movement recordings during reasoning? In W. Becker, H. Deubel, & T. Mergner (Eds.). *Current oculomotor research: Physiological and psychological aspects* (pp. 279-285). New York: Plenum Press.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle, G. (2000). Heuristic and analytic processes in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1713-1734.

Sclingensiepen, K. H., Campbell, F. W., Legge, G. E., & Walker, T. D. (1986). The importance of eye movements in the analysis of simple patterns. *Vision Research, 26,* 1111-1117.

Simon, H. A. (1982). *Models of bounded rationality.* Cambridge, MA: MIT Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3-22.

Sperber, D., & Girotto, V. (2002). Use or misuse of the Wason selection task? A rejoinder to Fiddick, Cosmides and Tooby. *Cognition, 85,* 277-290.

Sperber, D., & Girotto, V. (2003). Does the selection task detect cheater detection? In K, Sterelny & J. Fitness (Eds.), *From mating to mentality: Evaluating evolutionary psychology.* New York: Psychology Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition.* Oxford: Blackwell Publishing.

Stenning, K., & van Lambalgen, M. (2002). Semantics as a foundation for psychology: A case study Wason's selection task. *Journal of Logic, Language and Information, 10,* 273-317.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Lawrence Erlbaum Associates Ltd.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin.* Chicago: University of Chicago Press.

Stanovich, K. E., & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking and Reasoning, 4,* 193-230.

Stanovich, K. E. & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23,* 645-726.

Trimble, E. J. (1990). Report on accident to Boeing 737-400, G-OBME. Aircraft Accident Report No: 4/90, (EW/C1095).

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology, 57(1),* 165-182.

Van Duyne, P. C. (1973). A short note on Evans' criticism of reasoning experiments and his matching response hypothesis. *Cognition, 2*, 239-242.

Van Duyne, P. C. (1974). Realism and linguistic complexity. *British Journal of Psychology, 65,* 59-67.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology, Vol. I.* Harmondsworth: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20,* 273-281.

Wason, P. C. (1969). Regression in reasoning? *British Journal of Psychology, 60*, 471-480.

Wason, P. C. & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition, 3,* 141-154.

Wason, P. C. & Green, D. (1984). Reasoning and mental representation. *Quarterly Journal of Experimental Psychology, 36A,* 597-610.

Wason, P. C. & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluation of information in an inferential task. *British Journal of Psychology, 61,* 509-515.

Wason, P. C. & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23,* 63-71.

Weiner, P.P. (Ed). (1958). *Values in a universe of chance: Selected writings of Charles S. Peirce (1839-1914).* Garden City, NY: Doubleday.

Wittgenstein, L. (1953). *Philosophical investigations. Translated by G.E.M. Anscombe.* New York: Macmillian.

Yachanin, S. A. (1983). *Cognitive short-circuit strategies: The path of least resistance in inferential reasoning.* Unpublished PhD thesis, Bowling Green State University, Ohio.

Yachanin, S. A. & Tweney, R. D. (1982). The effect of thematic content on cognitive strategies in the four-card selection task. Bulletin of the *Psychonomic Society, 19,* 87-90.

Yama, H. (2001). Matching versus optimal data selection in the Wason selection task. *Thinking and Reasoning, 7,* 295-311.