TRAINING AND DUAL PROCESSES IN HUMAN THINKING

by

HELEN LOUISE NEILENS

A thesis submitted to the University of Plymouth
in partial fulfilment of the degree of

DOCTOR OF PHILOSOPHY

School of Psychology
Faculty of Science

June 2005

Training and Dual Processes in Human Thinking

Helen Louise Neilens

Abstract

The aim of the research presented in this thesis was to investigate the effects of training on reasoning and decision making performance. In Experiment 1 a study is reported which examined the relationships between performance on a variety of reasoning tasks and measures of individual differences. Tasks employed were documented in the literature for their differential responding according to heuristic and analytic processes. The reasoning tasks to be utilised in the training studies were also validated. In Chapter 4, two statistical training studies are reported which demonstrate that analytic responding on everyday reasoning problems can be increased after instruction on the Law Of Large Numbers. Bias was eliminated, but only on written justifications of their responses. Belief-based responding was still utilised when participants were asked for a quick indication of argument strength on a rating scale. This demonstrates a dissociation between analytic and belief-based responding. A second series of experiments explored the effects of both abstract and schema-based training on selection task responding. All the training procedures resulted in positive transfer apart from training on the logic of the material conditional which facilitated performance on arbitrary tasks only. Relationships between performance on the tasks post-training and cognitive ability indicated that training was more effective for higher ability participants. The differential training effects were discussed in terms of complexity of training procedures. The findings overall have implications for dual process theories of reasoning. The findings suggest that the interaction between training and System 1 and System 2 tasks/responses is a great deal more complicated than the simple analysis that is afforded by dual process accounts.

LIST OF CONTENTS

Chapter 1: A Review of Pragmatic and Analytic Processes in Reasoning and Decision Making

Chapter 2: Training and Transfer in Reasoning and Decision Making

Chapter 3: Individual Differences in Reasoning and Bias

CHAPTER 4: Effects of Training in Statistical Principles on Reasoning and Bias

CHAPTER 5: The Effects of Training on the Selection Task

CHAPTER 6: General Discussion

LIST OF TABLES

CHAPTER 4

Experiment 2

Experiment 3

CHAPTER 5

Experiment 4

Experiment 5

Experiment 6

LIST OF EXAMPLES

LIST OF FIGURES

List of Appendices

Appendix A – Materials and Instructions Employed in Experiments 1 – 6

APPENDIX A1 – Experiment 1

APPENDIX A2 - Experiment 2

Appendix B – Statistical Tables of Statistical Analyses: t-tests and ANOVAS for Experiments 1 – 6

Acknowledgements

Firstly I would like to thank my Director of Studies, Dr Simon Handley and my second supervisor, Professor Steve Newstead, who have been the best team that I could have wished for. The research in this thesis has benefited considerably from their knowledge. Their support, advice and encouragement has provided much motivation.

Secondly I would like to acknowledge the Economic and Social Research Council who provided me with the financial support without which I would not have been able to carry out this research.

I would also like to thank my family, my husband Phil and my children Finley and Tallulah, who have been without holidays and family weekends for 7 months whilst patiently waiting for me to finish this thesis. They make this all worthwhile.

A big thank you also goes to all my friends and family who have left me alone whilst I have been writing, and been very patient and understanding. Lastly I would like to say thank you to Paul Klaczynski, Keith Stanovich and Steven Sloman for advice and information in relation to the studies reported in this thesis.

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award except the permitted Postgraduate Diploma in Psychological Research Methods.

This study was financed with the aid of a studentship from the Economic and Social Research Council (Award No. R00429934481).

This thesis is the result of the author's own investigation, carried out under the guidance of her supervisory team. Other sources are explicitly referenced.

During this period of study, the author has also completed the Postgraduate Diploma in Psychological Research Methods, and achieved the following conference and seminar contributions.

Refereed Conference:

Neilens, H.L., Handley, S.J. and Newstead, S.E. (2004). The *Effects of Statistical Training on Reasoning and Bias.* Paper presented at the XXI BPS Cognitive Section Annual Conference, Leeds, 6-8 September 2004.

Seminar Presentation:

An extended version of the above conference paper was presented to the Thinking and Reasoning Research group at the University of Plymouth, March 2004.

Word Count: 81833

Signed. H. L. Neilens

Date. 23rd June 2005

CHAPTER 1

Pragmatic and Analytic Processes in Reasoning and Decision Making

1.1 General Introduction

1.1.1 Introduction to the review

The purpose of the research presented in this thesis is to investigate the extent to which heuristic and analytic processes affect performance in human thinking, using instruction and training procedures designed to improve reasoning competence. Two series of experiments will be reported in this thesis, which explore the effects of explicit instruction on human thinking, using tasks well documented in the literature for their elicitation of non-logical and biased responses: law of large numbers reasoning and the selection task. Of particular interest in this research are the effects the instruction procedures have on biased responding on these tasks and the extent to which training in one domain may generalise to other reasoning and decision-making tasks.

There has been an explosion of reasoning research over the last 30 years and the framework used now to test and interpret findings has changed substantially from when it began. This review will commence with a brief history of research in reasoning and decision-making, to provide the theoretical and conceptual background for the current research.

This will be followed by a discussion of a contempory theory of thinking, the Dual Process Theory, the development of which has materialized out of research findings related to bias and non-logical responding on a variety of reasoning and decision-making tasks. Supporters of Dual Process Theories propose that the brain consists of two separate systems, one related to analytic thought and the other intuitive. It is argued that logical and non-logical responses on reasoning tasks are the result of the different reasoning processes within the two systems, the analytic system facilitating logical reasoning and the intuitive system cueing non-logical or biased responses. Evidence for this view will be discussed in the context of logical and non-logical performance on the reasoning and decision-making tasks to be used in the experiments reported in this thesis.

1. 1. 2 A brief history of the deduction paradigm

Historically, research in reasoning has been conducted using the deduction paradigm that originated from the philosophical and psychological tradition of logicism. The proposal was that logic formed the basis for rational thought (Henle, 1962; In Evans, 2002). Piaget, a very influential researcher, proposed that according to the theory of formal operations, adults develop abstract reasoning based on an inherent mental logic (Inhelder & Piaget, 1958). The deduction paradigm was therefore used to demonstrate the extent to which people reasoned logically and became the basis for testing the rationality of people's reasoning in psychological laboratories (Evans, 2004; 2002).

Deductive reasoning is based on logical arguments and one of the most common paradigms used for studying deductive reasoning competence is syllogistic reasoning. Syllogistic reasoning was first devised by Aristotle and believed to be the basis of all rational thought (Evans, Newstead, & Byrne, 1993). People are presented with a set of premises and asked if the conclusion logically follows. See Example 1.1 that illustrates a syllogism consisting of two premises and a conclusion.

Some A are B,

No B are C,

Therefore, some A are not C.

Example 1.1

The premises must be assumed to be true and reasoning must be based on the premises only. Knowledge and prior belief is irrelevant to the task. A logically valid argument is one in which the conclusion can be shown necessarily to follow from its premises. In other words, if all the premises are true then a valid conclusion must be true.

Evidence from real life situations and in experimental laboratory settings demonstrate how people do not always reason in accordance with some normative system of logic. When participants are asked to perform reasoning and decision-making tasks many systematic errors and biases are produced. Bias has been defined as the systematic influence of some logically irrelevant feature of the task (Evans, 2002; 2004). It is consistently found that people ignore relevant information or attend to irrelevant features of the task. Performance is highly dependent upon the precise content and context in which a logical problem is presented (Evans, 1993).

People were judged as irrational and illogical due to the errors and biases that were elicited when performing on logical reasoning tasks (e.g. Wason, 1968; In Evans, 2002). It was argued that there was something wrong with people rather than the logic with which they were being compared. However these errors and biases are now known to be evidence of pragmatic, belief-based or heuristic and other nonlogical causes of responses to tasks. Pragmatics are defined as 'aspects of meaning and language use that are dependent on the speaker, the addressee and other features of the context and utterance' (Levinson, 1983). In other words these are the situational influences of the problem or argument itself. Beliefs on the other hand are personal and can be defined as any cognitive content held as true to the individual. In any given reasoning or judgement situation these two factors interact and inferences are drawn from both, which may or may not result in a biased response. Researchers in human thinking are taking into account such pragmatic influences and belief based responses in order to gain further understanding of the reasoning processes themselves (Cheng & Holyoak; Evans & Over, 1996; Evans & Over, 2004; Stanovich, 1999).

At the same time that people were being judged as irrational when compared to a normative logic in reasoning, a highly influential body of research was being conducted in decision-making which drew the same conclusions. Early research using the heuristics and biases approach (Tversky & Kahneman, 1971) concentrated on how incorrect statistical reasoning was performed. A variety of studies illustrated how intuitive responses resulted in errors and biases which led to the identification of two distinct approaches for responding to statistical problems. One was spontaneous, intuitive, effortless and fast, whilst the other was deliberate, rule-governed, effortful and slow. It was deemed not necessary to study correct statistical reasoning as controlled reasoning which led to correct answers was seen as a default case that needed no explaining (Kahneman & Frederick, 2002). However it is now deemed extremely necessary to study correct and incorrect responding on the tasks in order to observe the interactions between the two systems and in contrast to Tversky and Kahneman, it is the intuitive system that is seen as the default system.

1. 1. 3 Belief bias in human thinking: the case of syllogistic reasoning

A useful, illustrative example of a deductive reasoning task that consistently elicits biased responding is the syllogism involving realistic content. In abstract form syllogisms are a sound measure of deductive reasoning competence. However when realistic content is included people find it extremely difficult to dissociate the logic from their beliefs resulting in errors. It is a well established finding that people are more likely to accept the conclusion to a syllogism if they believe it, irrespective of its logical validity (Evans, Barston, & Pollard, 1983; Newstead, Pollard, & Allen, 1992). This has been termed the belief bias effect.

In this paradigm, participants are presented with four types of syllogism, Valid-Believable, Valid-Unbelievable, Invalid-Believable and Invalid-Unbelievable and they are asked to judge the validity of the conclusions (See Table 1.1 for an example of each taken from Evans, Barston, and Pollard (1983) plus the percentage of acceptance rates of conclusions as valid).

Example	Acceptance
Valid-believable	
No police dogs are vicious.	
Some highly trained dogs are vicious.	
Therefore, some highly trained dogs are not police dogs.	89%
Valid-unbelievable	
No nutritional things are inexpensive.	
Some vitamin tablets are inexpensive.	
Therefore, some vitamin tablets are not nutritional.	56%
Invalid-believable	
No addictive things are inexpensive.	
Some cigarettes are inexpensive.	
Therefore, some addictive things are not cigarettes.	71%
Invalid-unbelievable	
No millionaires are hard workers.	
Some rich people are hard workers.	
Therefore, some millionaires are not rich people.	10%

Table 1.1 Example syllogisms and acceptance rates of conclusions as valid taken from Evans et al. (1983)

Three basic effects have emerged from studies involving manipulation of belief and logic. Firstly, logically valid conclusions are more readily accepted than invalid ones, which is in line with the logic of the syllogism. Secondly, believable conclusions are more readily accepted than unbelievable ones which illustrates the strong effect of belief and thirdly there is an interaction between logical validity and believability. In other words, the effects of believability are stronger on syllogisms leading to invalid conclusions than on those leading to valid conclusions (Evans, Newstead, & Byrne, 1993).

The findings in the belief-bias literature provide evidence of irrational reasoning behaviour. Participants are able to follow the logic by endorsing more valid than invalid conclusions but they cannot ignore the believability of the conclusions either. To account for these findings, Evans and Over (1996; Evans, 1993; Evans, Over, & Manktelow, 1993) developed a theory of rationality based on two kinds of rationality, a personal and an impersonal approach. The personal rationality, termed Rationality1, incorporates the goals of the individual and whether the reasoning or behaviour is usually reliable for achieving those goals. The impersonal rationality, termed Rationality2, incorporates whether a normative system of logic such as propositional logic is being followed. Therefore it is possible to appear irrational according to some normative theory of reasoning but actually be completely rational in order to attain some goal or desire. In the real world people rely on prior knowledge and experience to achieve a practical goal or make a choice and therefore be rational1. In turn this may result in logical errors being made and appear irrational2. As Evans and Over (1996) pointed out, in the belief bias literature people cannot fully follow the instructions to reason logically and ignore prior belief. This indicates that people's reasoning does not fully involve a conscious process which is within their control. This rationality is highly bounded by cognitive constraints.

Evans, Over, and Manktelow (1993) interpreted the belief bias effect using the rationality1 approach. They proposed that mechanisms of reasoning have evolved to facilitate the achievement of goals in the real world rather than to solve problems in the laboratory and that this type of reasoning is highly adaptive. People have large belief systems which they use and draw on in order to help them achieve their goals. Conclusions that are consistent with pre-existing beliefs are not examined thoroughly because it would be advantageous to maintain beliefs unless there is a good reason to change them and because to question every current belief would be cognitively costly. However when the evidence contradicts a pre-existing belief people scrutinize it and refute it if possible as to accept it would introduce a contradiction and disrupt the internal consistency of their belief system. Thus for unbelievable conclusions people appear more rational. Another related explanation of the belief bias effect in terms of rationality1 is that of confirmation bias which is the tendency to seek out evidence which confirms rather than disconfirms current beliefs.

Stanovich (1999) also distinguishes between two systems of rationality. However his argument impacts directly on the arguments of evolutionary psychologists in the field of reasoning; evolutionary adaption and instrumental rationality. For the former, he suggested that adaptive optimization is at the level of the genes and for the latter maximisation is at the level of the individual. He proposes that an individual's information processing that is adaptively optimal could deviate substantially from a normative model because cognition is optimally adapted in an evolutionary sense. Stanovich also proposed that individuals of high analytic intelligence may respond in accord with normative rationality but those of lower analytic intelligence might be more likely to track evolutionary rationality in situations that put the two types of rationality in conflict.

Due to findings in the literature in relation to the influences of belief and logic on reasoning and decision making performance, several researchers developed a two-factor theory as an attempt to explain the co-existence of reasoning biases with deductive competence (Evans, 1984; Evans & Over, 1996; Stanovich, 1999; Sloman, 1996). The theory developed on the basis that the pragmatic factors which influence reasoning competence appeared to be automatic and preconscious. In contrast, people are able to verbalise the processes they utilise in analytic or deductive reasoning, with no awareness of underlying causes of nonlogical responses (Evans, 2002). In the next section dual process accounts of reasoning will be reviewed.

1.2 Dual Process Theories of Reasoning

1. 2. 1 Introduction to Dual Process Theories

In recent years, many researchers have adopted a two-process model of reasoning to explain the heuristic and belief-based influences on reasoning and decision making (Epstein, 1994; Evans & Over, 1996; Evans; 2003; Sloman, 1996; Stanovich, 1999). The hypothesis is that there are two distinct systems underlying reasoning and decision making. System 1 consists of both innate and domain-specific knowledge acquired through learning and System 2 is related to intelligence and analytic reasoning. It is believed that belief based and pragmatically cued responses can be interpreted under System 1 processing and logical or analytic responses can be interpreted under System 2. The two systems compete with each other for control dependent on the context of the problem or argument. Problem content can be manipulated to elicit different types of responses and instruction techniques have been shown to reduce pragmatic inferences. The experimental evidence for these accounts will be reviewed before a discussion of the tasks to be used in the following training studies.

Dual-process theorists argue that there are two separate cognitive systems underlying thinking and reasoning with distinct evolutionary histories (Evans, 2003). The systems have been termed implicit/explicit (Reber, 1993), associative/rule-based (Sloman, 1996; 2002), heuristic/analytic (Evans, 1984;1989), tacit/explicit (Evans & Over, 1996), rational/experiential (Epstein, 1994) and System 1/System 2 (Stanovich, 1999). There are differentiating factors between the dual-process theories that have been proposed but there are also several main features in which they all agree. System 1 processes (using Stanovich's (1999) terminology) are rapid, automatic, preconscious, heuristic-based and relatively undemanding of computational capacity. System 2 is controlled, conscious, analytic and related to working memory and measures of general intelligence. System 1 is comprised of a set of autonomous subsystems which include both innate input modules and domain-specific knowledge acquired through learning and experience whilst System 2 is thought to have evolved uniquely to humans. According to Stanovich, it has evolved as a 'long-leash' system with little genetic control, allowing humans to pursue their own individual goals rather than to act as slavish vehicles of the genes. It is argued that it is System 2 that is sensitive to instruction and permits abstract hypothetical thinking that cannot be achieved by System 1 (Evans, 2003).

A great deal of evidence in the reasoning and decision-making literature points to the existence of these two systems. Take for example the belief bias effects found in syllogistic reasoning. This task demonstrates how both logic and belief influence responding on a deductive reasoning task. Logical performance is attributed to System 2's analytic reasoning processes and the belief-based responses reflect System 1's intuitive responses. Individuals who are able to respond more logically to the task have been found to be higher in cognitive ability (Stanovich & West, 1997; Newstead, Handley, Harley, Wright, & Farrelly, 2004), which is in accord with the view that general intelligence is associated with System 2 thinking. Instructional procedures also seem to have an effect on the belief bias effects (Newstead et al., 1992) but we will return to this finding in more detail in Chapter 2. Studies of belief bias using neuropsychological techniques have also demonstrated that different areas of the brain are utilised dependent on the content of the task. Goel and Dolan (2003) showed that belief-laden and belief-neutral content in syllogistic reasoning elicited the use of different parts of the brain.

1. 2. 2 Dual process accounts of reasoning

How do individual proponents of dual process theories view the characteristics of the two systems? Sloman's associative system's (System 1) computations reflect similarity and temporal structure. He proposes the system encodes and processes statistical regularities of its environment, frequencies and correlations amongst the various features of the world. His rule-based system (System 2) consists of rules which are abstractions that apply to any statements that have a certain well-specified symbolic structure involving logical content (Sloman, 2002). Sloman posits that the signal that both systems are operating in a reasoning situation is that of simultaneous contradictory belief. This is when a problem causes individuals to believe two contradictory responses at the same time. Tversky and Kahneman's (1983) example of the conjunction fallacy, Linda-the-bank-teller problem is a good illustration of this. Participants are presented with a paragraph describing a hypothetical person, Linda:

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

Participants are then asked to rank order eight statements involving Linda, according to the statement's probability. One of these statements was 'Linda is a bank teller' and another 'Linda is a bank teller and is active in the feminist movement'. Tversky and Kahneman found that above 80% of participants rated the second statement as most probable, and yet these were graduate and medical students who had received statistical training. The associative system cues the response based on a judgement of similarity, or representativeness in terms of Tversky and Kahneman. The statement about Linda describes a person who sounds like a feminist bank-teller and therefore it is easy to imagine she is one rather than just a normal bank-teller. Of course logically it is more probable that Linda is a normal bank-teller because the probability of Linda being a bank-teller and active in the feminist movement cannot be higher than the probability of Linda just being a bank teller.

In Evans and Over's (1996) dual process model, individuals attain rationality by using the two systems. They distinguish between two different types of rationality, Rationality 1 which views an individual's cognitive processes as rational if they are useful for achieving their goals and Rationality 2 which proposes that people are rational if they are reason or make judgements in accord with a normative system such as standard propositional logic. Evans and Over propose that the tacit system affects the extent to which a person is rational 1 and the explicit system primarily affects the extent to which a person is rational 2. Much of their theory stems from Evans' (1984; 1989) original heuristic-analytic framework which views the implicit system as being responsible for the relevance processes utilised on reasoning and decision-making problems. These are automatic and habitual but it has been well illustrated that people can make conscious decisions although not always optimal due to the cognitive constraints of the explicit system. The heuristic-analytic framework will be discussed in more detail in a later section when we will consider different explanations in relation to performance patterns on the selection task.

According to Stanovich (1999) when faced with any problem System 1 processes are highly contextualized, personalized and socialized. System 2 processes in contrast serve to decontextualize and depersonalize problems as this system is better at utilising rules and reasoning analytically. Stanovich argues that many information-processing behaviours are carried out automatically in System 1 and due to the pervasiveness of the primacy effects of automatic contextualisation of problems in this system this has been termed the *fundamental computational bias*. Stanovich proposes that differences in cognitive ability are found only on problems that strongly engage both reasoning systems and in which the reasoning systems cue opposite responses, which incorporates most of the reasoning tasks to be discussed in this chapter. Likewise on the Linda-the-bank teller problem discussed above, it is the higher ability individuals who can override the System 1 inferences that lead to the conjunction fallacy and reason according to the conjunction rule of probability theory (p(A) >= p(A and B)).

In Stanovich's view it's not just cognitive ability but an individual's goals and beliefs that have an influence in achieving rationality. But how are these measured? One method is by using self-report measures of thinking dispositions. Stanovich (1999) proposed that cognitive capacities and thinking dispositions are constructs at different levels of analysis in cognition. Thinking dispositions are at the intentional level of analysis and incorporate individuals' epistemic goals whilst cognitive capacities refer to the computational limitations of an individual. Thus, the rational level of analysis concerns the goals of the system, the beliefs relevant to those goals, and the choice of action that is rational given the system's goals and beliefs.

Thinking dispositions are viewed as cognitive styles that are more malleable than cognitive ability (Stanovich, 1999) and therefore more teachable. Perkins, Jay, and Tishman (1993) for example identify seven broad thinking dispositions which may characterise good thinking i.e. to clarify and seek understanding, to be planful and strategic and to seek and evaluate reasons. Stanovich proposes that it is crucial to determine the relative proportion of variance in reasoning tasks that may be explained by both thinking dispositions and cognitive ability as the extent to which thinking dispositions explain variance in a rational thinking skill independent of cognitive capacity would predict that the skill would be more teachable.

Although the two systems are discussed by many reasoning researchers, only one places them within the context of a global theory of personality: Cognitive-Experiential Self-Theory (Epstein, 1994; Denes-Raj & Epstein, 1994; Epstein, Pacini, Denes-Raj, & Heier, 1996; Pacini & Epstein, 1999). Cognitive-Experiential Self-Theory is different from other approaches such as Stanovich's as it is about individual differences in thinking dispositions not ability. Epstein and his associates propose that there are two systems, the rational and the experiential, that operate in an independent, parallel and interactive manner and together contribute to behaviour with relative contributions varying from none at all to complete dominance by either one of the modes. Pacini and Epstein (1999) report associations between their self-report measure of Rational-Experiential information-processing and other personality measures. Rational individuals were more emotionally well adjusted, held a positive view of their self and the world, were able to exert self-control and control of events, could delay gratification and assume responsibility, be more flexible in their thinking style and had liberal values. Experiential individuals related well to others, communicated emotions readily, and were more tolerant, trusting, spontaneous and open-minded.

The Rational-Experiential Inventory is an orthogonal measure therefore individuals may be high on both scales or just one. Klaczynski, Gordon, and Fauth (1997) found that more rational individuals as measured by the REI produced more critical reasoning responses and fewer reasoning biases than relatively intuitive individuals. They propose that when information presented to people is inconsistent with their beliefs, then the rational system is engaged leading to more analytic reasoning strategies but when the information presented is consistent with their beliefs, then the experiential system is engaged leading to memory-based or heuristic responding.

One difference between Klaczynski's and Stanovich's dual process accounts is in relation to System 2. One of the functions of System 2 is proposed to be analytic reasoning. In order to achieve high level abstract reasoning, a person must evaluate arguments and evidence in a way that is not contaminated by prior beliefs and knowledge. This ability to decontextualise from both the contextual cues of the problem and prior knowledge is viewed as a necessary skill for high level analytic reasoning strategies but it is not clear in the literature what factors facilitate this skill.

Stanovich and West (1997) suggest that even if humans are optimally adapted to their environments at the rational level of analysis, there may still be computational limitations at the algorithmic level that prevent the full realization of the optimal model. They proposed that individuals with greater algorithmic capacity should show an enhanced ability to reason independently of prior belief. These individuals are also able to flexibly contextualise and decontextualise the problem dependent on what is necessary to resolve it. It is proposed that this ability is one of the key aspects of critical thinking.

In contrast, Klaczynski (Klaczynski, Gordon, & Fauth, 1997) argued that whilst cognitive ability predicts the sophistication of reasoning responses, thinking dispositions predict bias. Hence decontextualised reasoning is governed by thinking dispositions. In the next section of this review, the issues and conflicting views about which factors facilitate decontextualised reasoning will be discussed in relation to research performed by Stanovich (Stanovich & West, 1997) and Klaczynski (Klaczynski, Gordon, & Fauth, 1997).

1. 2. 3 Decontextualised reasoning and decision making

Stanovich and West (1997) designed an analytic technique, called the Argument Evaluation Task, for developing separate indices of a person's reliance on the quality of an argument and on their prior personal beliefs about the issue in question. Three hundred and forty-nine participants were assessed on their beliefs concerning 23 real social and political issues and then asked to evaluate arguments put forward by a fictitious person, on the same issues. For example given the statement 'It is more dangerous to travel by air than by car', participants had to indicate their degree of agreement or disagreement on a scale of 1 to 5.

Then in a separate testing session participants were given the following instructions followed by the test items:

Instructions: We are interested in your ability to evaluate counter-arguments. First, you will be presented with a belief held by an individual named Dale. Following this, you will be presented with Dale's premise or justification for holding this particular belief. A Critic will then offer a counter-argument to Dale's justification for the belief. (Assume that the Critic's statement is factually correct.) Finally, Dale will offer a rebuttal to the Critic's counter-argument. (Assume that Dale's rebuttal is also factually correct.) You are to evaluate the strength of Dale's rebuttal to the Critic's counter-argument, regardless of your feeling about the original belief or Dale's premise.

Dale's belief: It is more dangerous to travel by air than by car.

Dale's premise or justification for belief: It is more dangerous to travel by air than by car because air accidents are more likely to involve fatalities.

Critic's counter-argument: Passengers are 3 times more likely to be killed per mile travelled in a car as compared to a plane (assume statement factually correct).

Dale's rebuttal to Critic's counter-argument: Because reckless or drunk drivers cause the great majority of all automobile accidents (assume statement factually correct), car travel is at least safer than air travel for people who wear safety belts and travel with sober and careful drivers.

Example 1. 2

Participants were then asked to indicate the strength of Dale's rebuttal to the Critic's counter-argument from a scale of 1 (very weak) to 5 (very strong). Individual scores were compared to an objective argument quality index to assess how biased they were in their responses. The authors found that both cognitive ability and thinking dispositions were unique predictors of the ability to evaluate arguments independently of prior beliefs. In fact their unique variance as predictors exceeded the variance that they had in common with each other which emphasizes the distinction between the two factors.

According to Stanovich (1999) the ability to decontextualise on a variety of reasoning tasks is linked indicating that it is a domain-general dispositional trait related to the ability to reason independently of prior beliefs. He argues that normative responding on reasoning tasks requires some form of decontextualisation. In syllogistic reasoning involving realistic content, participants must decontextualise from their knowledge of the world or beliefs and on his Argument Evaluation Task participants must separate their prior opinion on the issue from evaluation of the argument. Stanovich and West (1998a; Stanovich & West, 1998b) conducted a series of studies to investigate the relationships between cognitive ability and thinking dispositions measures and performance on a variety of deductive and inductive reasoning tasks such as syllogistic reasoning, selection tasks, statistical reasoning and his argument evaluation task. Firstly they reported that all the tasks correlate with each other which indicates that those participants who respond normatively on one type of task tend to respond normatively on the other. Secondly they found that cognitive ability and thinking dispositions were predictive of reasoning performance on all the tasks. Cognitive ability was the strongest predictor. They showed that algorithmic limitations explained some performance discrepancies but thinking dispositions were able to predict individual differences on the tasks even after cognitive ability was partialed out (Stanovich & West, 1998a).

An opposing argument is put forward by Klaczynski, Gordon, and Fauth (1997). They agree that a crucial component of critical thinking is the ability to decontextualise one's reasoning from one's beliefs and goals but they propose that it is 'a metacognitive competence not captured by intelligence tests'. Klaczynski et al. found that biases in reasoning were best predicted by information-processing style (rational vs. experiential), with rational individuals being more likely to have a less biased reasoning style. This is in contrast with the view of Stanovich and West (1997) who believe that cognitive capacity governs your ability to decontextualise. In their account, individuals with a larger cognitive capacity can override the implicit/intuitive system and therefore be less biased. Klaczynski, Gordon, and Fauth found that general ability measures, such as verbal ability were dissociated from measures of information processing style. The amount of statistical reasoning that an individual engaged in was predicted by cognitive ability but the degree to which individuals were biased by belief was not. Thinking style governed bias independent of cognitive ability.

Newstead, Handley, Harley, Wright, and Farrelly (2004) consistently found that ability was related to logical performance in syllogistic reasoning. Individuals of higher ability are more able to resist the response cued by believability. They also found that rationality as measured by Epstein's Rational-Experiential Inventory was consistently associated with performance on non-conflict syllogisms i.e. ones where belief and logic were not in conflict, valid believable and invalid unbelievable problems. Newstead et al. suggest that the self-report measure of rationality is really measuring a participant's willingness to take part in the thinking tasks and the correlations may reflect the fact that people who score highly on this scale are taking the tasks more seriously than others. When the problems contain no conflict the cues for the correct response are obvious therefore those who take the task seriously perform well. They add that when the problems involve conflict such as valid unbelievable and invalid believable ones, then motivation is not enough to respond correctly.

It appears in the literature that there is agreement that decontextualisation is necessary on a variety of reasoning tasks in order to achieve optimum rationality but there is no clear consensus as to what facilitates that ability or skill. In summary, Stanovich proposes it is a skill dependent on ability and thinking dispositions. In contrast, Klaczynski argues that cognitive ability predicts the ability to perform on a task but thinking dispositions predict the amount of bias that will affect the performance independently. By using individual differences measures and instructional procedures in the research presented as part of this thesis, the aim is to investigate the extent to which they mediate the decontextualisation of prior knowledge and key elements from the context of the problem set. In turn this will provide more understanding of what personal and situational factors are required for optimum behaviour.

Although there are variations between researchers on the processes involved in the two systems described under dual process accounts of reasoning, the one commonality between them is the view that human thinking can be best understood as resulting from the operation of two distinct cognitive systems. In the next section we will review two types of reasoning task that have been the focus of considerable work by dual process theorists, law of large numbers reasoning and the selection task. These tasks will also be the focus of the training studies reported in Chapters 4 and 5. We will briefly review experimental work using these tasks and then consider dual process accounts of performance on them.

1.3 Everyday Critical Reasoning

1.3.1 Belief-based influences in everyday critical reasoning

Critical thinking has been defined as 'the ability and willingness to assess claims and make objective judgements on the basis of well-supported reasons' (Wade, 1995). This definition is advantageous in that it includes motivation as well as cognitive ability as being jointly important for effective reasoning and decision making. In other words it is necessary to be motivated as well as intelligent in order to self-consciously reflect on what is being thought about, deliberately examine the issues involved, ask why and think about reasons. In critical reasoning, people intentionally search for explanations and rules and make an effort to apply the rules to what they do. When people speculate, study, examine, experiment, ask questions, investigate, and/or find fault, they reason critically. This can be applied to all the types of reasoning and decision making tasks discussed in this thesis. Everyday reasoning problems refer to reasoning and decision making tasks which involve the evaluation of problems consisting of realistic content.

There is a great deal of evidence to suggest that belief-based influences impact on everyday reasoning, both in life in general and in previous research (Holland, Holyoak, Nisbett, & Thagard, 1986). Every day you can see many so called intelligent people who argue for their own point of view relentlessly, no matter what arguments are thrown back at them. It is argued that these departures from logical arguments are due to prior knowledge and biases which impact on reasoning and decision making behaviours. Nisbett and Ross (1980, Chapter 8) describe research findings as far back as the 17th century illustrating people's tendency to cling to preconceived beliefs and theories in the face of new evidence that should discredit them.

Lord, Ross, and Lepper (1979) asked participants to indicate whether or not they believed capital punishment to be a deterrent to potential murderers and then presented them with two hypothetical studies on the deterrent effects of capital punishment, one which supported their position and one which opposed it. Lord et al. found that participants rated the study that supported their position as 'more convincing' and 'better conducted' than the study that opposed their position. Participants treated supportive evidence carefully, whilst opposing evidence was analysed deeply. Participants were asked about their beliefs after reading only one of the studies that may or may not have supported their position. Lord et al. found that prior belief was strengthened if the study was consistent with their view but hardly affected if it was inconsistent with their view. After reading both studies, participants were more convinced by their initial belief than before they read them. Nisbett and Ross suggest that this work shows that people's responses to new evidence may be at times inappropriate.

Lord et al.'s study does seem to illustrate that different reasoning levels are utilised dependent on whether the evidence presented is favourable or not with a person's prior beliefs. Supportive evidence serves to strengthen a person's initial belief whereas opposing evidence using the same methods does not much affect belief. Confidence in a view is reinforced when presented with mixed evidence consisting of equal support for both positions indicating that people select the evidence that supports their own point of view.

The question to be asked at this point is why do people persevere in maintaining beliefs and theories? An automatic response to that is because they want to. Many beliefs and theories are strongly held therefore when challenged individuals will use whatever resources are available to discredit the contradictory evidence and bolster their own beliefs. However this is purely a cursory explanation and doesn't explain any of the underlying cognitive processes involved in belief-perseverance. One that may shed light on part of it is the confirmation bias explanation put forward earlier to explain the belief bias effect in syllogistic reasoning. In other words when people are confronted with a problem, they search for evidence that confirms prior beliefs or hypotheses over disconfirming evidence. Evidence to support the presence of a confirmation bias is illustrated in the confirmatory bias model of Koriat, Lichtenstein, and Fischhoff (1980) who studied decision making in experts. They proposed that people's inclination to use reasons from memory that confirm the focal hypothesis leads to a general tendency toward overconfidence. The stronger and more numerous the reasons that are employed, the greater the confidence expressed in the selected answer. However, because people overlook reasons against the selected answer, they are likely to be overconfident that they are correct.

To account for the higher scrutiny of evidence on belief opposing studies in Lord et al.'s experiment, Evans et al.'s selective scrutiny theory of belief bias in syllogistic reasoning may be applicable (Evans et al., 1983). For further discussion see Evans (1989). Individuals focus on the conclusion and if it is consistent with their beliefs they will then accept it but if it in inconsistent with their beliefs then they will examine the logic of the problem to see whether it is valid or not. On critical reasoning problems they will then use whatever cognitive resources are available to discredit the evidence. Whichever theory appears to be the most appropriate, one clear finding from all the above research is that people generally fail to appreciate intuitively the scientific strategy of disconfirmation or falsification therefore leading to the perseverance of incorrect beliefs.

More recently, Klaczynski and his colleagues have conducted a series of studies investigating individual differences in reasoning and bias on problems involving everyday content (Klaczynski & Gordon, 1996; Klaczynski, 1997; Klaczynski & Fauth, 1997; Klaczynski, Gordon, & Fauth, 1997). Using a wide variety of everyday reasoning tasks, they have consistently found that when presented with a conclusion that is incongruent with an individual's beliefs, more sophisticated reasoning strategies are utilised in order to discredit the evidence when compared to belief congruent information. One question to be asked was whether people suppress reasoning on belief congruent problems rather than engage in deeper processing on threatening problems. Klaczynski, Gordon, and Fauth (1997) addressed this by including problems involving conclusions which were neutral to an individual's goals and found people's responding on these to be as cursory as on belief congruent problems. This led to the conclusion that people are using deeper processing strategies on belief threatening problems.

Klaczynski et al.'s findings can be interpreted using a dual process account. To explain the different levels of reasoning strategies utilised dependent on the content of the problems Klaczynski and Fauth (1997; Klaczynski, 1997; Klaczynski & Gordon, 1996) proposed a model of self-serving reasoning. They suggested that because belief congruent evidence is consistent with their goals, then it is readily assimilated to individuals' belief systems and processed with relatively little effort. Therefore only less sophisticated reasoning strategies and task related memories are used. This is reflective of System 1's intuitive information processing. However, when the evidence is incongruent or threatening to an individual's goals, dissonance is created and the individual is compelled to expend considerable effort processing the evidence. The additional cognitive expenditure activates more sophisticated reasoning strategies that are normally difficult to access but which effectively reduce the dissonance. This in contrast is reflective of System 2's analytic reasoning strategies. The interaction of the two systems leads to optimum rationality in terms of achieving the goals of the individual, or in terms of Evans and Over (1996), attaining rationality1.

But what about decontextualised reasoning? As discussed before it is believed that it is a crucial ability for critical thinking tasks yet it is not being utilised on these problems. Klaczynski (1997) suggests that motivation may affect reasoning on everyday problem solving due to the individual involvement on a problem which may relate to the amount and type of cognitive resources allocated to formulating a solution. Motivational goals may distort the objectivity of information processing and lead to skewed reasoning. In other words thinking dispositions may predict the bias. For instance, Klaczynski, Gordon, and Fauth (1997) posit that individuals possess rough versions of sophisticated critical thinking rules but they only display their competence in using them when they are activated by a limited range of contextual conditions. One of these critical thinking rules is the law of large numbers.

1. 3. 2 Belief-based influences on law of large numbers reasoning

The law of large numbers is proposed to be an intuitive version of a statistical rule which people use to solve inferential problems in everyday life. The law of large numbers rule states, 'the certainty with which an inference about a population can be drawn increases as the size of a sample drawn from that population increases' (Klaczynski, Gordon, & Fauth, 1997). It is suggested that these rules are cued by elements of the problem. Nisbett, Krantz, Jepson, and Kunda (1983) state three different properties of a problem that makes statistical reasoning more likely. Firstly, the sample space and the sampling process are clear in order for it to be easier to see what information is relevant, for example it is clear that there is a single trial or many repeated trials as with throwing a dice. Secondly, the role of chance in producing events is clear and thirdly the culture specifies statistical reasoning as normative for the events. Nisbett et al. propose that these three elements operate at an individual level but often simultaneously to increase people's use of statistical heuristics on problems that require statistical thinking.

Previous research has provided evidence that people do use statistical concepts in solving problems in certain domains. Jepson, Krantz, and Nisbett (1983) gave participants problems involving probabilistic, objective and subjective content. Probabilistic problems involved content whereby participants had to draw conclusions about a population from sample data that clearly incorporated random variation. Randomness was made explicit by stating variation in sample outcomes, including a random generating device in the problem, or by stating that a sample was 'random' (see Example 1.3 for examples of each type of problem taken from Fong, Krantz, & Nisbett, 1986). Objective problems consisted of information where participants had to draw conclusions about characteristics of a population on the basis of 'objective' sample data but with no explicit information about randomness of the data. Subjective problems dealt with judgements about the properties of an object or person.

It was found that problems involving probabilistic content had evoked use of LLN reasoning the most. More than half the answers on objective domain problems were statistical but statistical answers for subjective domain problems were very low. This is surprising as many problems and arguments in real life are subjective. People are often found to overlook statistical variables such as sample size, correlation, and base rate when they solve inductive reasoning problems (Kahneman & Tversky, 1972).

Probabilistic:

Bert H. has a job checking the results of an X-ray scanner of pipeline welds in a pipe factory. Overall, the X-ray scanner shows that the welding machine makes a perfect weld about 80% of the time. Of 900 welds each day, usually about 680 to 740 are perfect. Bert has noticed that on some days, all of the first 10 welds were perfect. However, Bert has also noticed that on such days, the overall number of perfect welds is usually not much better for the day as a whole than on days when the first 10 welds show some imperfections.

Why do you suppose the number of perfect welds is usually not much better on days where the first batch of welds was perfect than on other days?

Objective:

A talent scout for a professional football team attends two local games with the intention of observing carefully the talent and skill of a particular player. The player looks generally excellent. He repeatedly makes tackles worthy of the best professional players. However, in one of the games, with his team behind by 2 goals, the player is fouled while attempting to score and has the opportunity to score on a penalty free kick. The player however misses by far. The other team then goes on to score another goal and therefore wins by 3 to nil.

The scout reports that the player in question "has excellent skills, and should be recruited. He has a tendency to misplay under extreme pressure, but this will probably disappear with more experience and better coaching."

Comment on the thinking embodied in the scout's opinion that the player (a) "excellent skills" and that the player has (b) "a tendency to misplay under extreme pressure." Does the thinking behind either conclusion have any weaknesses?

Subjective:

Two New Yorkers were discussing restaurants. Jane said to Ellen, "You know, most people seem to be crazy about Chinese food, but I'm not. I've been to about 20 different Chinese Restaurants, across the whole price range, and everything from bland Cantonese to spicy Szechwan and I'm really not very fond of any of it." "Oh," said Ellen, "don't jump to conclusions. I'll bet you've usually gone with a crowd of people, right?" "Yes," admitted Jane, "that's true, I usually go with half a dozen people or more from work," "Well, people that may be it!" said Ellen, "people usually go to Chinese restaurants with a crowd of people they hardly know. I know you, you're often tense and a little shy, and you're not likely to be able to relax and savour the food under those circumstances. Try going to a Chinese restaurant with just one good friend. I'll bet you'll like the food."

Comment on Ellen's reasoning. Do you think there is a good chance that if Jane went to a Chinese restaurant with one friend, she'd like the food? Why or why not?

Example 1.3

Klaczynski et al. (1997) investigated three critical reasoning competencies, the law of large numbers, the intuitive analysis of covariance and the ability to detect flaws in experimental designs. They ascertained participants' occupational goals in the first testing session and then constructed the reasoning problems for the second testing session to involve either goal-enhancing, neutral or goal-threatening evidence (see examples of goal-enhancing and goal-threatening conclusions in Example 1.4.). They found that reasoning competence could be predicted by cognitive ability but thinking dispositions were the best predictor of reasoning biases, as measured by the Rational Versus Experiential Inventory (RVEI, Epstein, Pacini, Denes-Raj, & Heier, 1995).

Goal-enhancing – The researchers concluded that the experience of being an accountant results in a greater sense of independence and of inner strength than does being an architect.

Goal-threatening – Dr. R. concluded that accountants are involved in more sexual harassment than are people who are members of other occupations.

Example 1.4

The RVEI is a 59-item questionnaire designed to assess the extent to which individuals rely on rational versus experiential information processing. Klaczynski et al proposed that rational processors are better able to reason beyond the boundaries set by their pre-existing theories or beliefs, thus are able to process goal enhancing information more objectively. In other words, they are able to reason more logically or analytically, without prior beliefs or knowledge affecting their judgment. In sum, decontextualised reasoning is predicted by thinking dispositions on these tasks. This is in opposition to Stanovich and West's (1997; Stanovich, 1999) argument that thinking dispositions and cognitive ability are both predictors of the ability to reason independently of beliefs.

A strong conclusion to be made from the findings in the everyday reasoning literature is that the type of rationality people seem to be achieving is Evans and Over's rationality1, rationality of purpose. Participants in these studies are making preconscious moment-to-moment changes in their response strategies dependent on the direction of belief in the argument. Some individuals are able to view all the problems more objectively but overall it is the automatic response to contextualise the problem or in Stanovich's terms the fundamental computational bias which prevails on belief-consistent evidence. When the evidence is contrary to beliefs then people are motivated to override the computational bias of System 1 leading to more analytic reasoning strategies.

1. 4 The Selection Task

1. 4. 1 Introduction to the selection task

The task to be reviewed in the next section is one that has been the most investigated by researchers in the field of reasoning since it was first reported by Wason (1966), the selection task. The selection task was designed as a measure of deductive reasoning competence but it has been argued by many that it is not a good method of measuring reasoning processes (Evans, 2002; Sperber & Girotto, 2002; Girotto, Kemmelmeier, Sperber, & van der Henst, 2001). However, for researchers examining heuristic and analytic reasoning and decision making processes, selection task presentations involving a variety of manipulations offer much insight on the different levels of information processing individuals have available for use.

Performance on the indicative selection task has been investigated widely as a method for studying deductive reasoning. The task set is one of hypothesis testing, although it requires deductive reasoning based on the logic of conditionals (Evans, Newstead, & Byrne, 1993). In the standard form of the task, participants are presented with four cards lying on a table and are informed that each card has a letter on one side and a number on the other. They are then given a conditional statement, 'If there is an A on one side of the card, then there is a 3 on the other side of the card' which corresponds to a conditional of the form 'if p then q'. On the top side of the cards, participants can see A, D, 3 and 7 which correspond to p, not p, q and not q respectively (see Example 1.5). Participants are then requested to choose the cards that need to be turned over in order to decide whether the statement is true or false.

Each of the boxes below represents a card lying on a table. Each one of the cards has a letter on one side and a number on the other side.

The rule is: "If a card has an A on its letter side, then it has an 3 on its number side."

As you can see, two of the cards are letter-side up, and two of the cards are number-side up. Your task is to decide which card or cards must be turned over in order to find out whether the rule is true or false.

A	D	3	7

Example 1.5

Under a normative analysis, the logically correct answer is to turn over the A and 7 cards as the statement can only be falsified by finding a card that has an A on one side but does not have a 3 on the other. In other words, the p and not q cards must be chosen. There is no point choosing the D card (not p) as it is irrelevant what is on the other side of this card, unless the rule is interpreted as a biconditional because the rule has no implications for cards with letters other than A. The 3 card (q) doesn't need to be chosen either as there may or may not be an A on the other side. The rule does not state that there must be an A on the other side of the 3 therefore the rule would not be contradicted.

In this abstract form, the majority of people get the selection task problem wrong. In fact only up to approximately 10% of the general population resolve it correctly (Evans, Newstead, & Byrne, 1993). The most common responses are the p card alone or the p and q cards (the two cards named in the rule). There have been a number of explanations put forward as to why there is such a low solution rate. Stanovich and West (1998a; 1998b) propose that it is only individuals of extremely high cognitive ability that are able to solve the task correctly. However, this view may be over simplified and will be discussed further later on in this chapter. Researchers in the field have used different procedural variations in an attempt to increase correct responding such as changing the instructions from testing whether the rule is true or false to seeing whether the rule has been violated (Griggs, 1984), an instruction commonly used in thematic versions of the selection task which will be discussed later in the chapter. Solution rates were not increased by this or by changing the wording of the statement such as 'Every p is a q' or altering the presentation of the cards to rule out the ambiguous reference to the 'other sides' of the cards (Evans & Over, 2004).

The review will continue with the standard form of the task, the indicative selection task, which will then be followed by a review of thematic forms of the task, namely the deontic selection task.

1. 4. 2 Pragmatic and analytic influences on the indicative task

As well as studying why such a low number of the general population resolve the indicative selection task correctly, it is just as interesting to reasoning researchers to investigate why the majority of people respond with a p only or p and q response. P/q responders have been found to be lower in ability than correct responders (Stanovich & West, 1998a). Wason's (1966) original explanation was that participants responding p/q were exhibiting a confirmation bias. In other words, participants were attempting to find evidence which conforms with the rule rather than disconfirms it (Evans, 1989). This proposal has now been rejected by most researchers in favour of the 'matching bias' explanation first put forward by Evans and Lynch (1973) who demonstrated that by including negations in the conditional rule 'if p then not q', participants choose the logically correct response. For instance, given the rule 'If there is an A on one side of the card, then there is not a 3 on the other side of the card', under a confirmation bias explanation, participants would now select the A and the 7 cards. However, if participants are 'matching' they should choose A and 3 as the rule will be falsified if they appear on the same card. Most participants when given the rule in the form 'if p then not q' select the matching cards p and q and therefore appear to solve the task correctly.

The typical responses and individual differences on the indicative selection task have prompted several important theoretical explanations in the reasoning literature. The first one to be reviewed here is Evans' (1984; 1989) heuristic-analytic account. Evans (1984) proposed a general theoretical framework within which to view pragmatic and analytic reasoning behaviours before the development of the modern dual process theory. He distinguished between two types of thought processes, heuristic and analytic. Heuristic processes (System 1) are preconscious and their function is to direct attention to problem information making it appear relevant. Information that is not deemed relevant is not processed any further. Relevant information is then subjected to analytic processing (System 2).

Evans argued that the selection task was a special case in which analytic processes played no role in the choice of cards that were determined purely by relevance. Analytic processes influenced reasoning on other tasks but they served only to rationalize choices on the selection task. In the case of matching bias on the indicative selection task, linguistically cued heuristic processing directs attention to the cards named in the rule. Evans termed these the if-heuristic and the not-heuristic, more recently, the matching-heuristic. The if-heuristic was based on the linguistic usage of *if* which is used to facilitate hypothetical thinking about a possible state of affairs specified by the antecedent. He proposed that the matching-heuristic reflects the use of negation in natural language, which is to deny presuppositions rather than to assert new information. Therefore, given the statements "the letter is A" and the "letter is not A", both are about the letter A and it is A that appears relevant. Therefore the letter (or number) referred to in the conditional forms the topic of discourse whether negated or not. Thus matching cards seem more relevant than mismatching cards (Evans & Over, 2004). Analytic processing is applied only to the information that is represented as relevant. However, a limitation of Evans' heuristic-analytic theory was that it did not specify how deductive processing, the analytic component, was achieved.

According to Stanovich (1999) it is only individuals with very high ability that are able to resist or suppress the strong linguistic cues to respond p/q. Following Evans' (1984; 1989) heuristic-analytic framework, Stanovich (1999) proposed that the p/q response on the abstract version is cued by preconscious linguistic heuristics. Participants who give this response have been found to be lower in ability than participants who give the correct p/not q response. In other words, individuals of lower ability are linguistically cued by System 1 to name the cards stated in the rule. P/not q responders on the other hand are assumed to be reasoning analytically. They have much higher ability scores which indicates that to inhibit the heuristic response an individual must have less computational limitations. System 2 processes can override the automatic cues of System 1 for individuals higher in cognitive ability. Stanovich argues that it is these individuals that can override the fundamental computational bias of System 1. It is proposed that these individuals of higher ability are able to decontextualise the key elements from the problem content.

More recently, Newstead, Handley, Harley, Wright and Farrelly (2004) demonstrated that when presented with more than one indicative selection task, low ability participants did not recognise that there were any similarities between the tasks and thus responded differently to them. More able participants were able to identify the similarities between the different problems and therefore respond the same on each, usually the items named in the rule. However, only the very high ability participants were able to resist invited inferences such as biconditional interpretations of the conditional rule and give the correct response.

Newstead et al. (2004) found in two out of their three studies that correct responders on the abstract selection task were not associated with higher ability scores. They suggest a two-stage theory of decontextualisation in order to attain correct responding on this task. Firstly the conditional rule must be decontextualised from the scenario which will lead to consistent responding, usually p/q. Then invited inferences i.e. the assumption that conditionals are biconditionals must be resisted. Newstead et al. found that participants in two of their studies only achieved the first level. These participants were of higher ability and responded consistently on the abstract task. Overall their participants were not high ability scorers but further evidence for their two stage model was provided in their final experiment. The participants were overall higher in ability than participants in the other two experiments and as would be predicted by Stanovich, individuals with the highest scores were more able to solve the task correctly. However, those who responded consistently on the abstract tasks after correct responders were partialed out were also higher in ability than inconsistent responders.

In contrast to the low percentage of correct responding on the indicative selection task, a version of the selection task that has consistently obtained higher correct response rates is the deontic version of the selection task. Correct solutions on the indicative task have consistently been related to intelligence, however ability differences on the deontic task have often been found to be attenuated (Stanovich & West, 1998a). This version of the selection task is rich in content and contextual cues which have been found to facilitate correct responding. A review of the findings related to how these pragmatic cues may affect responding will be given in the following section.

1.4.3 Pragmatic influences on the deontic tasks

Thematic versions of the selection task have been studied since the early 1970's. Originally it was thought that by including thematic or realistic content, facilitation of logical reasoning would occur. Wason and Shapiro (1971) reported that participants chose the p/not q cards 63% of the time after being presented with the conditional rule 'Every time I go to Manchester I travel by train'. At first this finding seemed to be reliable; however this was soon found not to be the case (Manktelow & Evans, 1979; Griggs & Cox, 1982). One finding that has been reliable involves the use of deontic conditionals. Deontic conditionals express rules to regulate or guide behaviour. They express obligation or permission statements about what 'must' be done or 'should' be done. A famous example in the literature is the Drinking Age Problem (Griggs & Cox, 1982). See Table 1.2 for the conditional rule and card choices. It has been found that most people choose the correct response of drinking beer and under age people corresponding to the p/not q cards.

Permission Rule	Cards
"if people are drinking beer then they must be over 18 years of age"	Drinking beer (p) Not drinking beer (not p) 21 years of age (q) 16 years of age (not q)

Table 1.2 The Drinking Age Rule and response choices

It could be argued that it is familiarity with the rule that facilitates reasoning but other deontic tasks have been used involving content that people are not familiar with that still result in high solution rates. For example, the Sears problem involves the rule 'if a purchase exceeds $30, the receipt must be approved by the departmental manager'. Participants are presented with the four options and most select p not q demonstrating facilitation effects even though participants are not familiar with the rule (Evans, 1989).

For deontic conditionals to facilitate reasoning there are a number of conditions that need to be met, including the presence of a scenario or context, familiarity with the conditional, its rationale as a rule for guiding behaviour is clear and an instruction to seek violators is included (Evans & Over, 2004). Unlike the findings with indicative selection tasks there is little difference in cognitive ability between solvers and non-solvers on the deontic selection task (Stanovich & West, 1998a) which suggests that analytic reasoning is not necessary on this type of task. Under Evans' (1984; 1989) heuristic-analytic framework, it is proposed that on deontic tasks card choices are still determined by relevance but instead of linguistically cued as with the indicative selection task, card choices are now pragmatically cued. Heuristic cues on the deontic task lead to the correct choices being made therefore analytic reasoning is not necessary for correct responding.

To account for the facilitated performance on certain thematic versions of the selection task, Cheng and Holyoak (1985) proposed the pragmatic reasoning schemas theory. These abstract schemas are domain-sensitive clusters of rules acquired from everyday-life experiences. Each schema consists of a set of production rules that will determine reasoning behaviour if the schema is triggered. Cheng and Holyoak argue that the 'permission' schema is relevant for the selection task – a situation in which some action 'A' may be taken only if some precondition 'B' is satisfied – and the permission schema fits most thematic selection tasks that have obtained reliable facilitation effects. They propose that the rules of the permission schema lead to the same conclusion as propositional logic therefore the correct answer is given if the schema is cued (Manktelow and Over (1991) argue that the rule is actually a conditional obligation). Cheng and Holyoak tested participants by presenting them with a selection task problem involving abstract content at the same time attempting to cue the permission schema. They found facilitation effects that had not been found previously on an abstract version of the task.

Support for pragmatic reasoning schemas has been provided by Griggs and Cox (1993) who replicated the previous results and found that factors that impact performance on the schema versions, for example, explicit negatives on the not-p and not-q cards do not influence performance on the standard abstract task. This was taken for evidence in support of pragmatic reasoning schemas. However Noveck and O'Brien (1996) demonstrated that the permission rule coupled with the abstract content was only partly responsible for the facilitation effects. Enriched task features were also responsible for the facilitation effects. They also tested the 'obligation' schema rule with abstract content but found it failed to facilitate task solution. Noveck and O'Brien argue that pragmatic reasoning schema theory involves only two domain-specific schemas so far, permission and obligation, and the permission schema appears more reliable but only when used in conjunction with other task features. Pollard and Evans (1987) also demonstrated that presenting the permission rule on the drinking age problem but omitting the minimal police officer scenario decreased the facilitation.

A second theory that addresses the question of which content feature facilitates performance on thematic versions of the selection task is Social Contract Theory (Cosmides, 1989). Rather than acquired knowledge as in pragmatic reasoning schemas Cosmides proposes that the understanding exhibited in deontic selection tasks is a product of innate thought processes termed Darwinian algorithms. The algorithms consist of domain-specific mechanisms and the original one proposed was for detecting 'cheaters' on social contracts. According to Cosmides, understanding of the rules of social exchange has been essential for human survival and people readily detect the possibility of cheaters as people are innately sensitive to them. Given the rule: 'if you take a benefit, then you pay a cost', Cosmides argues that everyone understands if you don't pay a cost then you are cheating. Her explanation for facilitation on deontic tasks is that the contents fit the benefit-cost structure in the cheater-detection module. It is argued that her theory is not general enough and does not explicitly distinguish between authorities who lay down the rules and the actors who follow them so as not to be called 'cheaters'. It also does not explain why people are so keen to identify cheaters in the first place (Manktelow & Over, 1991). Manktelow and Over claim that for a theory of conditional reasoning, subjective utilities and probabilities must be considered. Following this Manktelow, Sutherland and Over (1995) demonstrated that probabilistic factors and utility judgements play an important role in deontic thought and suggest that the underlying representations for this type of reasoning are best captured by mental models.

Oaksford and Chater (1994; 1996) criticise both schema theory and social contract theory for their domain-specificity and argue they are not general enough to account for reasoning performance on anything other than the deontic tasks. Oaksford and Chater propose that under a rational analysis, behaviour on the selection task can be viewed as optimizing the expected amount of information gained by turning each card. They see the selection task as an inductive not a deductive reasoning problem. The purpose of a rational analysis is to show that behaviour is optimally adapted to the environment and they demonstrate that the matching response (p/q) is because the expected information gain from the q card is greater than that from the non-q card. When solving the deontic version, participants choose cards to maximise expected utility.

The final account of reasoning on the selection task to be reviewed in this chapter is Sperber, Cara, and Girotto's (1995) theory of relevance. In contrast with the domain-specific theories discussed above, this is also a domain-general account of reasoning affected by content and context of the problem. Relevance in this theory is determined by two principles. Firstly information is more relevant if there is greater cognitive effect resulting from processing it, and secondly, it is less relevant if greater effort is required to process the information. Sperber et al. propose that there are three levels of increasing depth and effort by which people process conditionals. The first, level a, is simply Modus Ponens, there is always a q with a p. The second level b is when the conditional is read as meaning that there are some p cases and q cases and level c explicitly represents the conditional as ruling out cases of p and not-q. Level c requires the most cognitive effort and is the one attained the least on the indicative version. In fact, Sperber et al argue that the indicative version of the selection task usually does not induce representations deeper than level b therefore leading to the matching response p/q. Deontic versions on the other hand can induce deeper processing in which cases of p and not-q are represented as violations, especially when violation instructions are used. Pragmatic information cues participants to the correct response.

The notion that heuristic cues determine relevance is similar to Evans' (1984; 1989) heuristic-analytic framework. However Sperber et al. extend this further by arguing that participants' poor performance on the selection task is best explained by considering that the process of linguistic comprehension provides participants with intuitions of relevance which are highly content and context-dependent. Participants trust their intuitions of relevance and select cards accordingly (Girotto, Kemmelmeier, Sperber, & van der Henst, 2001). Sperber et al. also do not accept the distinction between heuristic and analytic processes and propose that all levels of representation and process in cognition are guided by relevance (Evans & Over, 1996).

Whichever theory is being supported, one of the clearest arguments is that the indicative selection task and the deontic task are two different problems, even though the correct response may be the same. Deontic terms such as 'must' or 'should' change the nature of the task into a decision making one and many researchers are approaching it as such i.e. Manktelow & Over (1991), Oaksford & Chater (1994), Evans & Over (1996). Evans and Over (1996; 2004) argue that deontic thought must be seen as an instance of hypothetical thought and related much more deeply to expected benefits and costs. Like Sperber, relevance still guides the processes in deontic reasoning but it is the benefits gained or the costs avoided which are deemed relevant to achieve the goals. For example given the conditional 'if you tidy your room, you may go to the party' a person would make a judgement based on the benefits to be gained from tidying the room. Deontic selection tasks include a scenario which informs the participant of the relevant authority that has set the rule e.g. in the Drinking age rule it is a policeman. The context of the scenario may suggest the costs to be avoided or the benefits to be gained by following the rule and participants are requested to find the violations of the rule. Most people are able to do this because of experience with rules in real life. Due to experience of such rules people respond correctly on deontic tasks without the need for much analytic reasoning.

Under a dual process account the correct response (p/not q) is made available by the operation of rapid pragmatic System 1 processes, without the need for conscious analytic reasoning. According to Stanovich & West (1998a) there are no differences in ability between solvers and non-solvers of the deontic selection task. Heuristic processors on the abstract version of the selection task remain heuristic processors on the deontic version, but are pragmatically cued (rather than linguistically cued as on the abstract version) to the correct response. Those who reasoned analytically on the abstract version (high ability) reason analytically on the deontic version and also reach the correct solution. Therefore any differences in ability are eliminated.

1.5 Conclusion for the Review

In Chapter 1, the review has illustrated how prior knowledge and pragmatics influence performance on a wide variety of reasoning tasks. By applying a dual process framework and investigating the roles of individual difference factors such as intelligence and thinking styles when responding on these tasks, we can see which factors mediate different types of responses.

It has also been demonstrated how important it is to decontextualise reasoning from beliefs in order to respond normatively compared to some logical system. Law of large numbers reasoning and correct responding on the indicative selection task are proposed to be related to System 2 under dual process accounts of reasoning. One way of encouraging the intervention of System 2 is by instruction or training, therefore in Chapter 2 a review of the training and transfer of reasoning skills literature will be presented, illustrating the valuable insights that training and instruction paradigms have to offer on reasoning processes utilised on a variety of tasks. By using training and instruction techniques on everyday reasoning problems and selection tasks it will be possible to investigate the effect of direct manipulation of System 2 analytic processes on belief-based System 1 responses.

Chapter 3 presents an individual differences experiment which was designed to investigate the relationships between performance on reasoning and decision-making tasks that have been discussed in this chapter, with the aim of identifying which tasks and measures to utilise in the subsequent studies. In Chapter 4 two experiments are reported which investigate the effects of statistical training, namely training in the law of large numbers, on critical reasoning performance and bias. The third experimental chapter, Chapter 5 presents three experiments aimed to investigate the effects of training in conditional logic and pragmatic reasoning schemas on conditional reasoning performance. Finally, Chapter 6 contains a general discussion in terms of the theoretical implications for the findings reported here and recommendations for future research.

CHAPTER 2

Training and Transfer in Reasoning and Decision Making

2. 1 General Introduction

2.1.1 Why training?

One of the principal aims of this thesis is to explore the effects of training and instruction on reasoning and decision making performance. The review in Chapter 1 illustrated that people are influenced in their reasoning by a whole range of non-logical factors. Several factors that affect optimum performance on everyday reasoning and deductive reasoning tasks were discussed. In sum, it is necessary to take into account people's goals and beliefs in relation to the context of the problem or argument as well as knowledge of the normative system that they may be using. In psychological experiments, people have often been judged as 'irrational' because they have not been reasoning in accord with some logical system. However, it is now recognised that the biases that are elicited may be due to knowledge and beliefs that the participants bring with them to the laboratory settings.

Dual process theories of reasoning were also discussed in Chapter 1. Under a dual process account there are two systems of reasoning. System 1 processes are pre-conscious, automatic, rapid, consist of domain-specific knowledge and only the final products are posted in consciousness. System 2 processes are thought to be distinctly human, conscious, analytic and tied to constructs such as working memory and intelligence. It is argued that performance on reasoning tasks demonstrates the co-existence of two anatomically distinct areas of the brain related to logic and belief competing for control dependent on the context of the problem.

It is generally agreed that the ability to decontextualise is a skill necessary for optimum responding on most of the tasks discussed. However there is disagreement in the literature as to the processes or constructs necessary for this reasoning ability. Stanovich and West (1997; Stanovich, 1999) proposed that the ability to decontextualise was related to intelligence and thinking dispositions. In contrast, Klaczynski argued that intelligence only predicted the sophistication of reasoning and thinking dispositions were related to the ability to decontextualise.

Our interest here is the extent to which training mediates the influence of heuristic and analytic reasoning processes. Let's return to the example of belief bias in syllogistic reasoning. Is there a way a person who is responding due to believability of the conclusion can be instructed or trained to be more objective, or to override those belief-based responses? In terms of dual processes, a person's responses are being cued by System 1's heuristic influences, but can they be instructed to utilise System 2's analytic processes? In other words can they be taught to utilise a logical or normative system, and ignore their prior beliefs and knowledge, i.e. to decontextualise?

It is proposed that System 2 is sensitive to explicit instruction (Evans & Over, 1996; Evans, 2003). However, the dual process systems are proposed to be independent therefore the training effects on System 2 may not impact on System 1 responses at all. There are questions that we need to answer, such as, if people can override or suppress responses cued by System 1, what abilities facilitate this i.e. intelligence, thinking dispositions, or both? By giving training will people make more use of their explicit analytic reasoning system and if so will their reasoning be less influenced by beliefs? And are there only certain individuals susceptible to training or instruction in the first place?

This review will start with a discussion of the findings in relation to instructional procedures utilised to reduce belief bias in syllogistic reasoning. These findings have been interpreted under a dual process account of reasoning. That is logical and belief-based responses being associated with System 2 and System 1 processes respectively.

2.1.2 The effects of instruction on belief bias in syllogistic reasoning

Findings in relation to the belief bias effect resulting from syllogistic reasoning involving thematic content have demonstrated that the manipulation of instruction leads to a reduction in belief-based responding (Newstead, Pollard, Evans, & Allen, 1992; Evans, Newstead, Allen, & Pollard, 1994). Newstead, Pollard, Evans, & Allen (1992, Experiment 5) compared responses between two groups of participants. One received a standard set of instructions and the other group a set of augmented instructions which stressed that participants should accept conclusions only if they necessarily followed from the premises, not if they were only possible conclusions (see Example 2.1 for complete instructional set).

Augmented Instructions

This experiment is designed to find out how people solve logical problems. In the booklet which you have been given there are 6 logical reasoning problems. Your task is to decide whether the conclusion given below each problem follows logically from the information given in that problem. You must assume that all the information which you are given is true; this is very important. If, and only if, you judge that a given conclusion logically follows from the information given you should write "YES" in the space below the conclusion on that page. If you think that the given conclusion does not necessarily follow from the information given you should write "NO". (Please note that according to the rules of deductive reasoning, you can only endorse a conclusion if it definitely follows from the information given. A conclusion that is merely possible, but not necessitated by the premises is not acceptable. Thus, if you judge that the information given is insufficient and you are not absolutely sure that the conclusion follows you must reject it and answer "NO".) Please take you time and be certain that you have the logically correct answer before stating it. If you have any questions, please ask them now as the experimenter cannot answer any questions once you have begun the experiment.

Please keep the instructions in front of you in case you need to refer to them later on. (REMEMBER, IF AND ONLY IF YOU JUDGE THAT A GIVEN CONCLUSION LOGICALLY FOLLOWS FROM THE INFORMATION GIVEN SHOULD YOU ANSWER "YES", OTHERWISE "NO".)

Please do not turn back and forth from one problem to another once you have started. You must not make notes or draw diagrams of any kind to help you in this task. Thank you very much for participating.

Example 2.1 Augmented instructions taken from Newstead et al. (1992, Experiment 5)

Newstead et al. found that the interaction between logic and belief disappeared under the instructions which stressed logical validity. This was followed up by Evans, Newstead, Allen, and Pollard (1994) who conducted a series of studies to investigate whether belief bias in syllogistic reasoning can be reduced or eliminated by verbal instruction in principles of reasoning. They found that the belief bias was maintained despite the use of the same augmented instructions which emphasised the principle of logical necessity. There was no impact on logical responding either. They concluded that the augmented instructions were not sufficient to suppress endorsement of fallacious conclusions unless the conclusions were apriori unbelievable.

Evans et al. (1994) noted that there was a much lower acceptance of invalid-believable conclusions in their study than in a previous one (Evans, Barston, & Pollard, 1983). Therefore, under the assumption that bias may be reduced on the invalid-believable conclusion by augmented instruction with the emphasis placed on the concept of necessity, albeit a weak finding, Evans et al. (1994) tried a different instructional manipulation. These instructions strongly emphasised the structure of the syllogism, explained the logical meaning of "SOME" and repeatedly emphasised the need to base responses only on the information given. The emphasis on the concept of logical necessity was omitted (see Example 2.2 for instructions). Evans et al. proposed that the effect of elaborating the instructions might be quite general, inducing the participant to adopt a more cautious attitude which in turn could be of interest in relation to debiasing in reasoning generally.

Evans et al. concluded that the bias may be reduced but not eliminated by the elaboration of logical principles and does not require a specific emphasis on the principle of logical necessity. In particular, they reported a reduction in the acceptance of invalid-believable arguments which they explained using a mental models approach. That is the search for counter-examples may be facilitated by elaborated verbal instructions. No differences were found on the logic indices after the elaborated instruction procedure.

Instructions

The experiment is designed to find out how people solve logical problems. In the booklet which you have been given there are 4 logical reasoning problems. You task is to decide whether or not the conclusion given does or does not logically follow from the information which is given above. The information takes the form of two statements (premises) which can be expressed symbolically as follows:

ALL B ARE A,
SOME C ARE B.

As you can see, the two premises tell us something about the relationship between three terms: A, B and C. The term B never appears in the conclusion, since the conclusion is a statement about the relationship between A and C, or vice versa. The conclusion to the above example is, therefore, "SOME A ARE C".

Since this is a problem requiring logical analysis, you should interpret the word "SOME" in its strictly logical sense; meaning AT LEAST ONE AND POSSIBLE ALL. So the statement "SOME B ARE C" does not necessarily also mean that "SOME B ARE NOT C".

In the booklet you will find four different logical problems. They are the same type of problem as the example problem which is shown above; however, the terms used will not be letters of the alphabet, but real words instead. Your task is to write down, below the conclusion given, "YES" if you judge that the conclusion necessarily follows from the information given, or "NO" if you judge that the conclusion does not necessarily follow from the information given.

You are reminded that you must base your decision on the information given in the two premises- and this information only. You must assume that all the information which you are given is true-this is very important. If and only if, you judge that a specific conclusion logically follows from the information given you should write "YES"; the conclusion given may not always be the correct one.

Please take your time and be certain that you have made the logically correct decision before stating it.

If you have any questions, please ask them now, as the experimenter cannot answer any questions once you have begun the experiment.

Please keep these instructions in front of you in case you need to refer to them later on.

REMEMBER, YOUR DECISION SHOULD BE BASED SOLELY UPON WHAT CAN BE DEDUCED WITH ABSOLUTE CERTAINTY FROM THE TWO PREMISES – AND THIS INFORMATION ONLY.

Please do not turn back and forth from one problem to another once you have started. You must not make notes or draw diagrams of any kind to aid you in this task.

Example 2.2 Instructions taken from Evans, Newstead, Allen, And Pollard (1994; Experiment 3)

These findings in relation to the effects of instruction on belief bias provide evidence that people have some degree of conscious System 2 control over their logical reasoning processes. People who normally respond to the believability of the evidence are able to override or inhibit the System 1 response to some extent and follow the logic of the argument instead. However, it must be emphasised that the bias was not eliminated but only reduced which demonstrates the pervasiveness of System 1's belief-based processing.

The effects of belief in syllogistic reasoning are in accord with performance on the everyday reasoning problems discussed in Chapter 1. When people are given a problem that is consistent with their beliefs they process the information at a very cursory level and accept the conclusion without searching for the flaws in the argument. Only when the evidence is inconsistent with prior knowledge do people engage in any analytic reasoning. One of the purposes of this research is to investigate whether the instructional effects found in syllogistic reasoning can be replicated on different types of problems where prior knowledge and beliefs may cue different responses.

The next section in this chapter will discuss training in critical thinking. First there will be a brief introduction to training in critical reasoning in general. This will be followed by a review of the training studies conducted to date in relation to statistical reasoning and conditional reasoning using the selection task which will provide the rationale for the training studies to be reported in this thesis.

2. 2 Training in Critical Thinking

2.2.1 General introduction to training in critical thinking

According to Nickerson, Perkins, and Smith (1985) it is possible to teach critical thinking as a general skill. Traditionally, a great deal of the literature on training in reasoning and critical thinking has evolved out of the educational domain and has concentrated on identifying different approaches to studying, in addition to training techniques designed to foster critical thinking performance on problems outside the domain of training (Wade, 1995; Wolfe, 1995; Doolittle, 1995; Bensley & Haynes, 1995; Jakoubek, 1995). Perkins and Grotzer (1997) reviewed several training interventions designed to improve 'intelligent behaviour' through direct teaching of reasoning and problem solving strategies or metacognitive strategies such as discussion of ideas and inductive reasoning. They posit that the key to successful interventions includes the consideration of strategies (problem solving, reasoning etc.), the monitoring and management of one's thinking or metacognition in other words, thinking dispositions and transfer. McMillan (1987) reviewed 27 different studies that investigated the effects of instructional methods, courses, programs and general experiences on changes in college students' critical thinking. What was clear from his review was that critical thinking does improve whilst attending college but there can be no firm conclusions about what factors affect the change and there is no consensus in the literature as to what specific measurements to make. It is necessary to obtain additional information about the development of critical thinking skills, such as the generality or domain specificity of skills, the influence of prior belief and individual differences in ability and skills. In turn, this will lead to researchers in the field of thinking and reasoning having more understanding of the cognitive processes involved.

Lehman and Nisbett (1990; Lehman, Lempert, & Nisbett, 1988) investigated the effects of different undergraduate courses on a variety of reasoning tasks, namely on statistical and methodological reasoning and on reasoning about problems in the logic of the conditional. Their research was motivated by the concept of formal discipline formulated by Plato. This is the notion that reasoning can be improved by teaching the rules within a particular domain which can then be generalised outside the bounds of that domain. In their original study, in support of the formal discipline position, Lehman et al. (1988) found that graduate training in the probabilistic sciences (medicine and psychology) improved statistical and methodological reasoning. Training in these and law (defined as a nonscience) improved conditional reasoning, and training in chemistry (a deterministic science) resulted in no improvement on any of the types of reasoning examined. They argued that the improvements were due to the rule systems taught by the different courses. In their longitudinal study, Lehman and Nisbett (1990) replicated most of the results with psychology and social sciences undergraduates having improved statistical and methodological reasoning whilst the natural sciences and humanities undergraduates showed some improvement on these but mostly on conditional reasoning. Undergraduate training in psychology and the social sciences resulted in no improvement in conditional reasoning. Lehman and Nisbett conclude in support of the Formalist view that highly general inferential rules can be taught and the quality of reasoning improved.

According to Bransford, Sherwood, Vye, and Rieser (1986) it is important that people have access to domain-specific knowledge and general metacognitive knowledge for effective transfer. They discuss the point that training studies that result in failure of transfer involve individuals being taught to use strategies but not how to understand why they were useful and when to use them (Brown, Campione, & Day, 1981). An emphasis on executive or metacognitive processes can result in an improvement in thinking. Consistent with this view, Stanovich (1999) suggested that unlike cognitive capacity which is fixed, thinking styles are malleable and therefore teachable.

In accord with Stanovich's proposal, Klaczynski and Gordon (1996) provided participants with instructions aimed to motivate them to be more accurate. They found that overall performance was increased but bias was not reduced. However, it does not matter how flexible or rational people are in their thinking style, or to what extent training or instruction procedures aim to foster better critical thinking styles, people still need to be provided with the tools with which to reason. In the following section, a discussion of the training studies conducted to date in relation to statistical reasoning, namely the Law of Large Numbers will be provided. These studies have been very effective in increasing the amount of statistical reasoning utilised on a variety of everyday reasoning problems.

2. 2. 2 Training and transfer effects in statistical reasoning

Research had been conducted looking into intuitive rule systems for a long time before the modern dual process theories that have been discussed so far were proposed. The law of large numbers is proposed to be an intuitive version of a statistical rule which people use to solve inferential problems in everyday life. The rule states, 'the certainty with which an inference about a population can be drawn increases as the size of a sample drawn from that population increases' (Klaczynski, Gordon, & Fauth, 1997). It is suggested that these rules are cued by elements of the problem such as explicit information on sampling (Nisbett, Krantz, Jepson and Kunda, 1983) and the more cues that are present in the problem the more likely people are to utilise statistical heuristics on problems that require statistical thinking.

Fong, Krantz, and Nisbett (1986) conducted a series of experiments looking specifically at the law of large numbers. They proposed that if this inferential rule system was abstract, then people's statistical reasoning on everyday events would be improved after training on the rule system. In turn, according to the formalist view of reasoning, use of the rule would transfer across to other domains. In a series of experiments, participants were either trained in the LLN rule system which consisted of a description of the concept of sampling and the law of large numbers; given examples training which consisted of three problems in a given domain followed by an explanation of how to solve the problem in LLN terms; or Full training which consisted of Rules training followed by Examples training (see Example 2.3 for the rules training script).

Rules Training

We are very interested in studying how people go about explaining and predicting events under conditions of very limited information about the events. It seems to us to be important to study how people explain and predict under these conditions because they occur very frequently in the real world. Indeed, we often have to make important decisions based on such explanations and predictions, either because there is too little time to get additional information or because it is simply unavailable.

Experts that study human inference have found that some common-sense principles are helpful in explaining and predicting events, especially under conditions of limited information. One such principle of probability that is particularly helpful is called the *Law of Large Numbers*. In this study, we will teach you the Law of Large Numbers by introducing you to the probabilistic terms and ideas associated with this principle, and then provide examples to illustrate how the Law of Large Numbers can be used to explain and predict events in the real world.

Imagine an urn that is filled with gumballs. Let's say that the urn contains a very large number of gumballs - thousands, millions, or larger. The gumballs in this urn are known collectively as the <u>population</u>.

Example 2.3 Rules Training continued:

Let's say that there are two types of gumballs in the urn - red gumballs and blue gumballs. When we do this, we can now say that the population has two categories or groups, namely - red gumballs and blue gumballs.

Now let's say that in this population of red and blue gumballs there are 70% blue gumballs and 30% red gumballs. If that is the case, then we know more than that the population has two groups (red and blue): we now know the proportion of the gumballs in the blue group (70%) and the proportion of the gumballs in the red group (30%). This is known as the population distribution (other examples of distributions are 60% red and 40% blue, or 85% blue and 15% red, etc., but in every distribution, the sum of the proportions must be 100%.

Let's summarise what we've covered so far:

- A population is the entire set of objects we are interested in (all of the gumballs in the urn).
- Groups refer to the types of objects in the population (red and white).
- Distribution refers to the proportion of objects in each group (70% white and 30% red in this example.

One of the major goals of statistics is to find out something about a population. More specifically, we want to find out what the population distribution is. One way that we might do this would be to actually examine all of the objects in the population and count up the number of objects in each group. In our example, we would empty the entire urn and count the number of red gumballs and the number of blue gumballs. Using this method, we could find out exactly what the population distribution of red and blue gumballs was. But, there is a very serious problem with counting all of the objects in the population: populations, in general, are very large. If we were to count all of the objects in our gumball population, it would take more time and effort than would be practical (imagine counting a million gumballs!)

OK - so counting the entire population is impractical. What do we do instead to find out what the population distribution is?

What we do instead is to take a sample of the population. A sample is a subset of the population. We can take a sample of any size - if we pick 5 gumballs, we say that the sample size is 5; if we take 60 gumballs for our sample, we say that the sample size is 60, and so on.

When we take a sample from the population, we will get a sample distribution. The sample distribution is the proportion of objects in each group for the sample, just as the population distribution is the proportion of objects in each group for the population. For example, if we take a sample of 10 gumballs, we might get 6 reds and 4 blues. In this case, our sample distribution would be 60% red and 40% blue. We also might have happened to get 9 blues and 1 red, in which case the sample distribution would be 90% blue and 10% red. The important point here is that samples are estimates of populations. Since it is often impractical or sometimes impossible to examine the entire population, we instead have to draw sample to estimate what the population is like.

Some samples will have sample distributions that are closer to the population distribution than others. For instance, in our gumball example, a sample of 9 reds and 1 blue would be a very poor estimate of the population, while a sample of 8 blues and 2 reds would be a pretty good estimate of the population. The critical question is: *What determines how likely it is that samples will give good estimates of the population?* The answer is simple: if the samples are chosen haphazardly, or randomly (by, for example, mixing the urn and reaching into the urn blindfolded and scooping out the needed number of gumballs OR by mixing the contents of a gumball machine and letting the gumballs out one by one), then there is only one factor – *sample size.*

This brings us to the Law of Large Numbers: as the size of a random sample increases, the sample distribution is more and more likely to get closer and closer to the population distribution. In other words, the larger the sample, the better it is as an estimate of the population.

Example 2.3 Fong, Krantz, And Nisbett's (1986) Rules training on the law of large numbers

After reading this description, a demonstration was performed by the experimenter using a large glass urn filled with blue and red gumballs. The script was designed to adhere closely to the description that had just been read to maximise participants' understanding of the concepts they had just read about. Each of the concepts introduced in the description was highlighted in the demonstration, for instance the population distribution of the gumballs in the urn was 70% blue and 30% red.

All the main concepts were reintroduced and then the experimenter drew four samples of size 1, then four of size 4 and then four of size 25, returning the gumballs to the urn after each one. Each sample was summarized on a blackboard, monitoring the deviation between each sample and the population. Participants were informed that in accordance with the law of large numbers, the average deviation of a sample from the population would decrease as the sample size increased. Therefore the larger samples taken would deviate less from the population than the smaller samples.

Participants in the full training condition were then presented with three examples one of which is illustrated in Example 2.4. Each example involved a problem consisting of objective information in which participants had to draw conclusions about characteristics of a population on the basis of objective sample data, which contained no explicit information about randomness of the data (Jepson, Krantz, & Nisbett, 1983). Participants were requested to consider each one before turning the page to read an analysis of it in terms of the law of large numbers.

Participants were then given everyday reasoning problems consisting of probabilistic, objective and subjective information as described in Chapter 1. Probabilistic problems involve content whereby participants had to draw conclusions about a population from sample data that clearly incorporated random variation and subjective problems dealt with judgements about the properties of an object or person.

Objective Problem

Susan is the artistic director for a ballet company. One of her jobs is auditioning and selecting new members of the company. She says the following of her experience: "Every year we hire 10-20 young people on a 1-year contract on the basis of their performance at the audition. Usually we're extremely excited about the potential of 2 or 3 of these young people - a young woman who does a brilliant series of turns or a young man who does several leaps that make you hold your breath. Unfortunately, most of these young people turn out to be only somewhat better than the rest. I believe many of these extraordinarily talented young people are frightened of success. They get into the company and see the tremendous effort and anxiety involved in becoming a star and they get cold feet. They'd rather lead a less demanding life as an ordinary member of the corps de ballet."

Comment on Susan's reasoning. Why do you suppose that Susan usually has to revise downward her opinion of dancers that she initially thought were brilliant?

Analysis

We can analyse this problem using the laws of large numbers by thinking of each ballet dancer as possessing a <u>population</u> of ballet movements. Susan is interested in excellence, so we can divide the members of each population into two categories: 'brilliant movements' and 'nonbrilliant, or other movements'. We can think of the <u>population distribution</u> as the percentage or proportion in each category. For many dancers, the population distribution is actually 0% brilliant and 100% other; these dancers simply lack the talent to perform brilliant movement. For many other dancers, there is a small or moderate percentage of 'brilliant movement' gumballs in their urn. A true ballet star would therefore have a population distribution with a greater percentage of 'brilliant' movements than an ordinary member of the corps de ballet.

By conducting auditions, Susan is observing <u>samples</u> of each dancer's population distribution. An audition, however, is a very small sample of a dancer's movements. We know from the law of large numbers that small samples are very unreliable estimates of the population. When a dancer performs some brilliant moves during an audition, it is often because the dancer has happened to draw a couple of the 'lucky gumballs' that day: it does not prove that the population distribution for that dancer consists of a large percentage of 'brilliant movements'. It is reasonable to think that there are really very few dancers that have population distributions with a large percentage of brilliant movements; and so when Susan sees a dancer performing brilliantly at audition, the chances are it is just a lucky draw from a dancer who is capable of performing some, but not necessarily a great number of 'brilliant movements'. Therefore, when Susan hires such dancers and evaluates them after seeing a much larger sample of their movements, it is not surprising that she finds that many of these dancers that were brilliant at audition turn out to be only somewhat better than the rest.

Example 2.4 Objective problem taken from Fong, Krantz, And Nisbett's Law of Large Numbers Training

Fong et al. found that both formal rule training and examples training improved statistical reasoning and enhanced the quality of the reasoning for problems across all three domains.

The full training was found to have an additional effect. The studies did indeed provide evidence for the domain-independence of training. Participants who were given examples only training using problems from the objective domain were able to utilise the statistical principles across all three problem domains on testing. Further evidence was provided in their second experiment when they gave examples training to their participants in one of the three domains, objective, subjective or probabilistic, and found that the training significantly increased use of statistical reasoning regardless of the domain of training. From the results it may be concluded that participants are able to map the LLN rules they have learnt onto a pre-existing set of abstract intuitive rules that may then be used on problems in different domains to the one that they have been taught. The fact that examples training alone resulted in no domain-specificity effects suggests that learning based on specific problem types may be abstracted to a degree sufficient for use on widely different problem types. These findings are consistent with a dual process account of reasoning on everyday problems. People can learn to follow rules explicitly and respond normatively (Evans & Over, 1996). In other words, people can learn to be more rational in terms of a normative system of reasoning.

Fong and Nisbett (1991) addressed a couple of questions that were raised following this series of studies. Firstly, it could be argued that the domains used in the original study were too broad and there were no domain-independence effects of training at all. Secondly, it was possible that details of the training examples are still held in a participant's memory for a short while after training. Fong and Nisbett added in a two week delay between training and testing and used more tightly defined domains. Participants received a short introduction to the LLN principle followed by three examples in the domain of either sports or ability testing. Each example was again followed by an answer in terms of LLN. Use of the statistical principles was still improved after a two week delay, however domain specificity of training was observed. There was a loss of training effects over the delay in the untrained domain but these participants still applied the LLN heuristic more than participants from the untrained Control group. Fong and Nisbett argued that examples training served in part to provide a set of coding rules that link the inferential rule system more tightly to events in the trained domain. The training should transfer to other domains but especially in those persons who have formed the coding rules. Kosonen and Winne (1995) provided support for these findings and the Formalist view that given effective instruction in abstract, formal rules of reasoning, students can transfer those rules across domains.

The question is can people transfer rules of conditional logic in the same way? Can people be instructed in the rules leading to use of those rules on subsequent problems consisting of different content? The next section will review the literature on training and transfer effects in conditional reasoning using the selection task.

2.2.3 Training and transfer effects on selection task reasoning

Cheng, Holyoak, Nisbett, & Oliver (1986) explored the question of whether people have intuitive rules for conditional logic as they seem to for the law of large numbers. Cheng et al. (1986, Exp.1) gave their participants logic training in the material conditional. The training was manipulated in that participants were given either training on abstract principles of standard logic, or examples of selection task problems involving realistic content followed by an explanation of the correct response, or both combined (see Appendix 2A for details of training). They found that the rule training alone was ineffective and they proposed that this was because individuals have no ability to apply it to concrete problems. They also found examples training alone to be ineffective and suggested that individuals have no intuitive grasp of the rule they are being shown how to apply. However, training in standard logic, when coupled with training on examples of selection task problems, led to improved performance on subsequent selection task problems involving both arbitrary and deontic content.

Cheng et al. proposed that abstract principles coupled with examples served to elucidate the mapping between abstract principles and concrete instances. They suggested that knowledge of abstract rules of logic and ability to apply them are two separate skills and typically college students have not yet acquired either of them. This is very different from Fong et al. who found that rules training alone and examples training alone were effective in increasing use of LLN, and the effects of both together were additive. Thus providing further evidence for the existence of intuitive heuristics for LLN reasoning, but not for the material conditional.

A problem with the findings from this experiment was that the examples used as part of the logic training were not abstract. They involved real life content which may have facilitated reasoning on subsequent thematic selection task problems. In fact one of the examples was deontic, the Sears problem. This is one of the confounds that will be addressed directly in Chapter 5 of this thesis.

It could be argued that one set of instructions was not enough to facilitate logical reasoning. Cheng et al. provided participants with a semester long logic course which covered topics in propositional logic. However no significant improvement was found in the number of problems solved correctly. Evans and Over (1996) reported that they were not surprised that the results were negative. They argue that the selection task was not a good tool for investigating whether training in logical principles and attendance at logic courses would facilitate general deductive reasoning competence. According to Evans' (1984; 1989) heuristic-analytic theory very few people would get past a relevance judgement on a selection task consisting of arbitrary content.

But is there a way that individuals can be encouraged to reason more explicitly on these problems? The evidence in the belief bias literature discussed before demonstrates that certain instructional manipulations can reduce the belief bias effects in syllogistic reasoning. It is possible that there is a way to facilitate analytic reasoning on arbitrary selection task problems, thus overriding or inhibiting the implicit responses. This is not relevant for deontic tasks as the correct response is cued pragmatically anyway; hence it is not necessary to override the response cued by System 1. However, it should be possible to facilitate the improvement of deontic reasoning also as the correct analytic response on the arbitrary task is the same as the correct one on the deontic task.

Cheng and Holyoak (1985) proposed that people reason using 'pragmatic reasoning schemas', clusters of rules that are acquired through experience. Each schema is abstract but domain-sensitive. Two examples of these are the 'permission' and 'obligation' schemas. As with LLN reasoning, it was proposed that it should be easy to evoke the use of pragmatic reasoning schemas by presenting individuals with problems consisting of semantic cues designed to trigger them.

In Cheng et al.'s Experiment 3 they investigated the hypothesis that if people normally solve problems using pragmatic reasoning schemas then it should be possible to improve people's deductive reasoning by training them on them. This would also then lend support to Fong et al.'s suggestion that abstract training in naturally occurring rule systems can be effective in encouraging people to use them. The obligation training consisted of details of the nature of obligations in abstract format and the procedures necessary for checking if a violation of the obligation has occurred. An example of an obligation statement presented in the *if-then* conditional form was given (see Example 2.5 for the full script).

Obligation Schema Training

As you know, an obligation arises whenever it is the case that certain circumstances or situations create an obligation to perform some action. Obligations can often be stated in an *"If. . . then"* form. For example, the following regulation specifies an obligation: "If a student is a psychology major, then the student must take an introductory psychology course." More generally, if we call the initial situation *I* and the action *C*, an obligation has the form, "If *I* arises, then *C* must be done." In our first example, *I* is " being a psychology major," and *C* is "taking an introductory psychology course."

In order to assess whether an obligation is being satisfied, we need to consider the four possible situations that might arise. These are

1. *I* occurs.
2. *I* doesn't occur.
3. *C* is done.
4. *C* is not done.

Corresponding to each of these possible situations is a rule related to the fulfilment of the obligation. These rules are the following:

1. If *I* occurs, then it is obligatory to do *C*. Clearly, if *I* arises then failure to take the required action would constitute a violation of the obligation. To use our example, if a student is a psychology major, then that student must take an introductory psychology course.
2. If *I* does not occur, then the obligation does not arise. Consequently, *C* need not be done, although the person may do *C* anyway. For example, if a student is not a psychology major the student is not obliged to take an introductory psychology course. It may be permissible, however, for an English major to take an introductory psychology course. But in any case, the basic obligation is simply irrelevant if the student is not a psychology major.
3. If *C* is done, then the obligation is certainly not violated, regardless of whether or not *I* has occurred. If *I* did occur, then the obligation is satisfied. If *I* didn't occur, then the obligation didn't even arise (Rule 2). For example, if we know a student has taken an introductory psychology course, we can be sure the obligation has not been violated: Either the student was a psychology major, and hence fulfilled the obligation, or the student was not a psychology major, in which case the obligation didn't arise.
4. If *C* has not been done, then *I* must not have occurred. This is because if *I* had occurred, then the failure to do *C* would constitute a violation of the obligation. Thus, if a student has not taken an introductory psychology course, the student must not be a psychology major, or else the obligation will have been violated.

If you understand the above four rules, you should find it very easy to assess whether or not an obligation is being met. Note that there are only two situations in which it is possible for an obligation to be violated: When *I* occurs (and *C* is not done) (Rule 1), and when *C* is not done (and *I* occurs) (Rule 4). In the other two situations the obligation can't be violated. These are the cases in which *I* doesn't occur (in which case the obligation doesn't arise) (Rule 2), and in which *C* is done (in which case the obligation will have been met if it arose) (Rule 3).

You may wish to reread these instructions carefully in order to be sure you understand the rules for evaluating obligations. You will then be able to apply what you learned to the test problems. The test problems will include both obligations and other similar types of regularities. You will find it easy to solve these problems if you carefully apply Rules 1-4.

Example 2.5 Obligation Schema training taken from Cheng, Holyoak, Nisbett, and Oliver (1986), Experiment 3

Participants were given four arbitrary problems and four obligation problems for the test. They reported that obligation schema training was effective in producing a better performance on obligation problems. Cheng et al. also found that the benefit of obligation training extended to two of the arbitrary problems. When they investigated this further they found that response patterns under each condition on two of the problems resembled those of obligation problems. They concluded that some participants might have interpreted the two arbitrary problems as obligation situations after training. The two that it is suggested were possibly interpreted as obligations were; 'if a bolt of cloth has any red threads in it, then it must be stamped with a triangle' and 'if a house was built before 1979, then it has a fireplace'. It could be argued that both these tasks are highly thematic, full of semantic content and much more likely to be understood as deontic conditionals rather than ones that are purely arbitrary in content.

Therefore, Cheng et al. proposed that they had found support for the pragmatic reasoning schema hypothesis on two counts. Firstly, participants made significantly fewer errors on obligation problems than on arbitrary problems. Secondly, training in a pragmatic reasoning schema encourages use of that schema for problems that compel a semantic interpretation consistent with the schema and also, may refine participants' understanding of situations that are potentially interpretable in terms of the schema. They also found that training improved performance on two arbitrary problems. Cheng et al. suggest that this is because the procedure used provided participants with information as to which cases constituted violations in addition to simply orienting them toward checking violations. As these are consistent with the material conditional then they can be applied to arbitrary problems also.

2.2.4 Cross domain transfer

Providing support for Cheng et al.'s findings, Berry (1983) found thematic to arbitrary transfer effects when participants were given minimal explanations of thematic problems in an initial set and were instructed to verbalise the reasoning behind their answers to each problem. Klaczynski, Gelfand, and Reese (1989) failed to replicate this finding when they conducted a series of experiments to observe the transfer effects between thematic and arbitrary selection task problems. Participants were presented with either five arbitrary or five thematic tasks initially. If they were in the explanation condition, they were read in-depth explanations of the correct and incorrect selections immediately after they completed each of the first five problems (see Example 2.6 for examples of arbitrary and thematic problems plus explanations). Performance was then measured on a set of four arbitrary tasks. Klaczynski et al. found that the explanations did facilitate transfer but only from arbitrary to arbitrary tasks. No transfer occurred from thematic to arbitrary tasks.

Arbitrary Example

Your task on the following problem is to determine whether or not this rule is in effect:

"If 1/2 is on one side of a card, then Helium must be on the other side of the card."

This rule may or may not be in effect.
The four cards you have in front of you each has two pieces of information on it. On one side of each card, there is a fraction, and on the other side of each card, there is a chemical element. For two of these cards, you can see the fraction on one side of the card, but you cannot see the chemical element on the other side of the card. Fro the other two cards, you can see the chemical element on one side of the card, but you cannot see the fraction on the other side of the card.
Your task is to indicate the card or cards you would need to turn over to see whether this rule is in effect For which of these four cards would the information on the other side of the card help you test whether or not this rule is in effect?
Be sure to select all those cards that you need in order to make your decision, but only those cards that you definitely need..

(a)	(b)	(c)	(d)
3/4	1/2	helium	oxygen

Explanation

The correct answer is to turn over both the '1/2' and the 'Oxygen' cards. The rule states that if there is a 1/2 on one side of a card, then Helium must be on the other side of that card.

The '1/2' card must be turned over because the truth of the rule is determined by the chemical element on the other side of this card. If the element on the other side of this card is any element other than helium, then the rule cannot be true. Likewise, the 'Oxygen' card must be turned over because, according to the rule, Oxygen cannot be on the other side of a card that has 1/2 on it. If a card with Oxygen on one side has 1/2 on the other side of it, the rule cannot be in effect. In order to see if the 'Oxygen' card has 1/2 on the other side of it, one must turn this card over.

Example 2.6 Arbitrary and Thematic training examples continued:

The '3/4' card does not have to be turned over because any chemical could be on the other side of this card and the rule could still be in effect; that is, the chemical that appears on the other side of this card does not affect the truth of the rule, so the information on the other side of this card is irrelevant. The 'Helium' card does not have to be turned over either. Because the rule is not reversible, a card with Helium on one side does not have to have 1/2 on the other side of it; either 1/2 or 3/4 could be on the other side of this card and the rule could still be in effect.

Thematic Example

Imagine you are a police officer, on duty, walking through a local bar. It is your job to ensure that the local drinking laws are in effect in this bar. One such law is:

"If a person is drinking beer, then that person must be at least 19 years old."

The rule may or may not be in effect.
The four cards you have in front of you each represents one of the people drinking at the bar. There are two pieces of information about a person on a card: a person's age is on one side of a card, and what the person is drinking is on the other side of the card. For two of the people, you can see their age but you cannot see what they are drinking. For the other two people, you can see what they are drinking, but you cannot see their age.
Your task is to decide what additional information you would need in order to decide whether or not the drinking law is in effect. For which of these four people would the information on the other side of the card help you decide whether or not the bar is observing the law?
Be sure to select all those cards that you need in order to make your decision, but only those cards that you definitely need.

(a)	(b)	(c)	(d)
Drinking beer	22 years old	Drinking coke	16 years old

Explanation

The correct answer is to turn over both the 'Drinking Beer' and the '16 years old' cards. The law states that is a person is drinking beer, then that person must be at least 19 years old. The 'drinking beer' card must be turned over because the truth of the law is determined by whether or not that person is over 19 years old. If the age on the other side of this card is under 19, then the law cannot be in effect. Likewise, the '16 years old' card must be turned over because, according to the law, this person cannot be drinking beer. If a person who

is 16 years old is drinking beer, then the law cannot be in effect. In order to see if the 16-year-old person is drinking beer, one must turn this card over.

The 'drinking coke' card does not have to be turned over because any age could be on the other side of this card and the law could still be in effect; that is, the age of a person drinking coke does not affect the truth of the law, so the information on the other side of this card is irrelevant. The '21 years old' card does not have to be turned over either. Because the law is not reversible, a person who is 21 does not have to be drinking beer; either drinking beer or drinking coke could be on the other side of this card and the law could still be in effect.

Example 2.6 Arbitrary and thematic selection task training examples followed by in-depth explanations of the correct and incorrect responses taken from Klaczynski, Gelfand and Reese (1989)

Klaczynski et al. developed a contextual similarity hypothesis to account for the effects. According to this, the context of a problem consists of related elements from past experience or from the problem itself, thus similar problems share elements. The greater the number of shared elements, the greater the similarity between problems. The greater the similarity the more likely there will be transfer between problems. It was proposed that the elements in the arbitrary problems only differ in the words used for p, not p, q and not q, therefore are very similar resulting in transfer. In contrast, the elements in the thematic problems are tied to past experiences and are dissimilar to each other. Transfer is less likely to occur from thematic to other thematic problems or to other arbitrary problems.

In their second experiment, Klaczynski et al. tested the contextual similarity hypothesis however it failed to provide any supporting evidence. Participants failed to transfer between similar thematic pairs as they did dissimilar ones. Klaczynski et al. replicated their first experiment but included thematic problems as well as arbitrary problems as their test items and found that transfer occurred within problem types i.e. arbitrary to arbitrary, thematic to thematic, but transfer between problem types was unidirectional. That is transfer occurred from arbitrary to thematic problems only, not the other way round. Support for these findings has been provided by Bassok and Holyoak (1989) who found training on algebra and physics problems transferred to problems within the same problem type but transfer across domains was again unidirectional. Transfer occurred only from algebra to physics, not the other way round.

Klaczynski et al. (1989; Klaczynski & Laipple, 1993) proposed that when given training in an abstract domain, the rules are relatively general because there is no concrete knowledge base onto which the rules can be mapped. Therefore, once the rule is induced, transfer may occur to different problem types that are perceived as instantiations of the general category of problems to which the rule applies. Training in a domain-specific rule however may lead to the rule being utilised on other problems from the same domain as one-to-one mapping of the elements between problems is more readily perceived. Transfer to problems in other domains does not occur as the rule is tied to a specific domain. In terms of Cheng et al.'s logic training, the rules are relatively general therefore should transfer to arbitrary and deontic problems. According to Klaczynski and Laipple, because the logic training is likely to be consistent with the rules of the deontic problems, the training may be directly mapped onto and activate the schemas which may then be applied to other problems. Training in a domain-specific schema however should result in activation of that schema which will facilitate the solution of further problems within that domain with little or no transfer to problems from other domains. However, Cheng et al. did find some transfer from their obligation schema training to arbitrary problems but no transfer from their logic training (when examples were excluded) which is in the opposite direction to Klaczynski's proposals.

It was proposed by Klaczynski (1993) that transfer of rules among permission and arbitrary problems may be viewed as an aspect of more general learning processes. He specified two types of learning. During the training phase participants may have used inductive reasoning to abstract common elements from the explanations and to construct the rules. However a different mechanism is involved for transfer. That is learning to apply previously acquired information to new domains. This involves noticing the similarities between the problems, retrieving the rule learned and mapping this onto the test items.

As part of this research, it is necessary to investigate the inconsistencies in the previous research. To summarise, Cheng et al. found transfer effects after obligation schema training but Klaczynski et al. argue that transfer did not occur from domain-specific training. Under a dual process account we could argue that explicit training impacts on

System 2 therefore training on a domain-specific System 1 process such as obligation schemas, will not transfer to other types of problems.

Is it an aspect of the training itself that is causing these discrepancies, or the items the participants are tested on? It may be both or neither. It is possible that individual differences in ability or cognitive style may provide the answer or at least some of it. In the following section a review of the individual differences research in relation to training and transfer of reasoning ability will be given. This may provide more insight into the processes involved in successful (and unsuccessful) reasoning transfer in addition to further evidence for dual process theories.

2.2.5 Individual differences and transfer

Klaczynski and Laipple (1993) investigated the role of intelligence when transferring domain-independent and domain-specific knowledge to target domain-independent and domain-specific problems. They reported strong relationships with ability when transferring knowledge from permission problems to other permission schema problems and it was proposed that individuals with higher ability were able to perceive the underlying structural similarities between different problems and therefore apply the same rules to each. It is these individuals that can decontextualise a problem from its thematic content therefore resulting in positive transfer. Klaczynski and Laipple found only weak relationships with ability when transferring from abstract problems. They proposed that the training leads to the induction of a domain-independent rule which may be more easily applied to target problems.

This appears inconsistent with Stanovich's (1999) proposal that it is the participants of higher ability that are able to engage in analytic reasoning on the indicative selection task. It is these individuals that are able to override or inhibit the heuristic response. However, in

Klaczynski and Laipple's study participants were not tested for ability and reasoning performance relationships prior to training which is the relationship that Stanovich proposes. Klaczynski and Laipple are proposing that irrespective of ability, participants acquire the domain-independent rule which they can then utilise on subsequent problems from other domains. In contrast, to acquire and transfer a domain-specific rule, participants must be of higher ability.

As discussed before, thinking styles may also have a role to play in receptiveness to training. Houdé and Moutier (1996; 1999) conducted a series of studies aimed to investigate the influence of training on heuristic responding directly. Based on Evans' (1989) proposal that heuristic processes are involved in pre-attentional selection of certain features of a problem that are deemed relevant by the participant, they posited that people are 'inefficient inhibitors' which is why they are susceptible to making the biased responses such as matching bias (p/q responding on the indicative selection task). Houdé and Moutier (1996) used an experimental procedure where participants learnt to inhibit the matching bias response on the selection task. Participants were presented with the Wason selection task in the letter and number form and asked to state which cards had to be turned over to verify the rule "If there is an A on one side of a card, then there is a 3 on the other." After responding the participant was given an analysis of the problem and the trap it involved in order to make the reason for their failure clear (see Example 2.7 for details of script).

They found that the inhibition training increased logically correct responding and decreased the matching bias response. However these findings were not robust as the effect disappeared when the participants were tested eight days after training. Houdé and Moutier also reported that over half the participants still failed to inhibit the matching response after training and suggested that there may be individual differences in receptiveness to the training.

Inhibition Training

'In this problem, the source of the error lies in a habit we all have of concentrating on cards with the letter or number mentioned in the rule (the experimenter points to cards A and 3, and to the place where they are mentioned in the rule) and not paying attention to the other cards. This can have a very *misleading* effect on us: we think this makes things easier when in fact *we're falling into a trap*! This is probably what you did.'

The experimenter goes on to say:

'Thus, the goal here is (1) to not *fall into the trap* of the two cards A and 3 mentioned in the rule, and (2) to consider all of the cards, A, D, 3, 7, one be one by imagining the number or the letter it might have on the other side to see whether these cards can make the rule false... To help you understand, let's consider the different answers and eliminate the wrong ones – the ones that make you *fall into the trap* – to find the right answer.'

The experimenter then goes through the responses A, D, 7, A-3, and A-7 informing the participant why they might be falling into the trap and what they need to do (i.e. inhibit the response to name the cards in the rule and imagine what is on the other side of the cards not named in the rule). A box is used by the experimenter to depict the repertoire of schemes that may be used. The different responses are represented on cards (A, A-3 and A-7) which can be slid into the box to illustrate the inhibition and the activation processes.

The training ended when the participant was able to produce correct explanations for wrong answers A and A-3. The experimenter concludes by saying;

'So in this problem, you should not let yourself be *misled* by the cards with a letter or number mentioned in the rule (the experimenter points first to cards A and 3, and then to the place where they are mentioned in the rule), which makes you neglect the other cards. You might think this makes things easier but in fact you *fall into a trap*!'

Example 2.7 Inhibition training details taken from Houdé and Moutier (1996)

Houdé and Moutier (1999) replicated their previous findings and reported that receptiveness to training was related to a field-independent cognitive style. That is an ability to perceive an element separate from its geometric content and to adopt an analytic attitude in problem-solving. Participants reported to have a field-dependent cognitive style were not receptive to the training. These findings are consistent with Stanovich's and Klaczynski's proposals that thinking styles are important when investigating reasoning and decision-making behaviours. They propose that it is people with more open-minded and rational thinking styles that are able to decontextualise elements from thematic content.

Going with the notion that thinking styles are malleable therefore more susceptible to training, then it is these individuals that are more receptive to training.

Houdé and Moutier (1999) also studied the functional neuroanatomy of the biased-to-logical shift in the same set of participants performing the same task after cognitive inhibition training. They conducted a Positron Emission Tomography (PET) scan study to determine what brain regions were activated before and after training. A microgenetic change in the cortical anatomy of reasoning shifted from the posterior part of the brain in the pre-test to a prefrontal network on the post-test. They concluded that inhibition is the cognitive mechanism that can redirect attention. Houdé and Moutier argue that these findings are interesting for two reasons. Firstly the brain regions activated on the pre-test when matching bias was in effect are known to be involved in the semantic processing of shapes and colours and in the visuo-spatial processing of information. This they propose corresponds to the dual origin of the perceptual matching bias – visuo-spatial and lexico-semantic: visuo-spatial selection of the two shapes mentioned in the rule. Secondly, the brain regions activated on the post-test belong to a prefrontal network involved in working memory. What is different about these studies compared to the others discussed so far is that the training did not involve repetition of the task or logical training but purely matching-bias inhibition.

2.3 Summary and Conclusions

2.3.1 Summary of findings in relation to training studies

The literature reviewed in this chapter illustrates several conflicting findings. Experience and training in an intuitive rule of reasoning such as the law of large numbers does result in positive transfer to other problems involving realistic content. However domain-specificity of training effects are found after a delay. Training in the material conditional does not transfer or even increase correct responding on the indicative selection task unless examples are included as part of the training procedure. However this could be due to the selection task not being an effective measure of deductive reasoning. People are guided in their responses by pre-conscious heuristics of relevance which results in biased responses. Evans (1989) proposed that few people get past a relevance judgement on the selection task, but in line with the findings in syllogistic reasoning, there must be a way to increase normative responding on this task.

Cheng et al.'s pragmatic reasoning schema training did appear to facilitate reasoning on tasks involving different content but it is necessary to explore this further due to the realistic test items that were used in their study. On the other hand, Klaczynski and associates reported that schema training only transferred to other problems in the same domain and only for individuals of high ability. It is possible that individual differences in thinking styles may predict people's susceptibility to such training as in Houde and Moutier's inhibition training. Or the ability to understand and transfer the training may be due to intelligence as proposed by Stanovich and Klaczynski. This research aims to explore all these aspects.

The following section will provide a rationale for the studies reported in this thesis. General predictions in relation to the training and tasks to be used will be made in terms of current dual process theories of thinking.

2.3.2 Rationale for training studies

Most of the training studies presented in this chapter have illustrated the positive effects of verbal instruction on reasoning processes. The effects of belief were reduced in syllogistic reasoning after the elaboration of logical principles, although logical responding was not increased; rule-based training in statistical reasoning led to an increased use of the rule on problems involving different content and structure; Cheng et al.'s logic training (full training including explanation of the rules plus examples) increased correct responding on subsequent selection tasks as did their obligation schema training; Houdé and Moutier's studies reduced the matching bias response on the selection task; and Klaczynski et al.'s examples training facilitated the increase of correct responding on different problems. Albeit that there are discrepancies and conflicting findings between all these studies, they do support the conjecture that instruction is effective in eliciting explicit thinking (Evans & Over, 1996; Evans, 2003).

One purpose of this research is to find the source of these discrepancies discussed so far and to consider in detail the predictions that arise from a dual process framework. The question is what would dual process theories predict? Dual process theorists agree that System 2 is sensitive to instruction but in what way does that affect the reasoning processes, in both systems? Surely that would depend to some extent on the type of reasoning being performed i.e. statistical or deductive. As discussed in Chapter 1 and here in this chapter, it has been proposed that the law of large numbers is an intuitive rule system that individuals acquire through experience in a domain-specific manner. People have been shown to use the rule when the context of the problem suggests it, or when motivated to do so in order to maintain beliefs. Thus people may be seen as rational1, or rational in terms of their individual goals or beliefs. Fong et al.'s statistical training studies provide evidence that people can learn to conform with a rule leading to an increase in their rationality2 (Evans & Over, 1996).

The research on the effects of statistical training has consistently shown that use of LLN reasoning can be increased on a variety of problems. What would happen if participants received everyday reasoning problems that involved evidence that leads to belief-motivated reasoning as in Klaczynski et al.'s studies, after they had received LLN training?

A series of studies will be reported in Chapter 4 which approaches this question directly. The previous research suggests that LLN reasoning would be increased but beliefs are, in dual process terms, a System 1 process. People respond automatically or analytically dependent on whether the information is consistent or inconsistent with their prior beliefs. The aim is to investigate the extent to which training affects responding on everyday reasoning problems designed to elicit belief-based responses. As discussed in Chapter 1, when presented with problems involving information that is unbelievable participants utilise a more sophisticated reasoning strategy to refute the evidence. However when the information is believable, people tend not to analyse the evidence too deeply. By giving participants statistical training the aim is to observe whether use of statistical reasoning will be increased, and whether the training will impact on bias. It is possible that the effects of belief are too strong to override. At the same time the studies aim to replicate the findings of Fong et al. (1986) that training in an intuitive rule system improves use of the rule for problems in the same domain and transfers to other problems involving different types of content. One prediction that falls out from the instruction effects on belief bias and from Evans and Over (1996) is that statistical reasoning will be improved, and the effect of belief may be reduced, but it will not be eliminated.

The selection task has long been viewed as a measure of deductive competence. People have been judged as irrational when compared to some logical system, for not applying analytic reasoning strategies resulting in the correct response on the arbitrary version of the task. Responses on both arbitrary and deontic forms of the selection task are thought to be cued by relevance, either linguistically on the arbitrary version or pragmatically on the deontic version. By looking at the task this way, it may be easier to judge whether explicit training has any effect on these processes and how, therefore resulting in increased rationality2.

In Chapter 5 a series of studies are reported which aim to clarify the effects of different training techniques on the selection task. There are discrepancies in the previous research that have been discussed in these two introductory chapters which are necessary to explore. Findings reported so far leave it unclear whether the conflicting results are due to the training procedures themselves, the test items used or individual difference factors such as intelligence or thinking dispositions. Both Cheng et al. and Klaczynski et al. found transfer effects after schema training although slightly conflicting ones. However their training was completely different. Cheng et al. found some transfer from obligation to arbitrary problems but Klaczynski et al. argue that the transfer between domains is only unidirectional, from arbitrary to thematic. The aim here is to resolve some of these conflicting results in the literature, but again what would be predicted under a dual process account?

Klaczynski et al. (1989; Klaczynski & Laipple, 1993) would predict that abstract training (both theirs and Cheng et al's logical rules) would transfer to problems involving either arbitrary or deontic content. There would be no relationship with ability as a general rule would be formed during training which would help in solving subsequent tasks from whichever domain. After domain-specific training though, transfer would only occur to other problems from the same domain and the individuals who are able to do so would be higher in ability. These individuals would be able to recognise the structural similarities resulting in one-to-one mapping of the elements.

Further predictions will be made in relation to the individual training experiments in the relevant chapters. As previously discussed, all the reasoning and judgement tasks to be used have been shown to elicit both logical and non-logical responses. Usually these tasks are investigated separately, either in the reasoning literature or the decision making literature. By exploring performance on all of them as part of this thesis it is possible to examine the relationships between them and at the same time determine the extent to which performance on each draws on System 2 resources.

In the following chapter, an individual differences study will be reported. Stanovich proposed that the ability to decontextualise on a variety of critical thinking tasks is linked (Stanovich & West, 1998b; Stanovich, 1999). It is necessary to investigate this proposal in relation to the tasks to be used in the training studies and in relation to measures of individual differences in cognitive ability and thinking styles. Performance on these tasks also provide evidence for the two processes of thinking and decision making as discussed in these two introductory chapters. The study was conducted in order to replicate the findings reported in the literature and to identify the tasks to be used in the following training experiments.

CHAPTER 3

Individual Differences in Reasoning and Bias

3.1 Introduction to Experiment 1

In Chapter 1 a review of the different reasoning tasks was presented, which illustrated how prior knowledge and beliefs can influence performance on a wide variety of reasoning problems well documented in the literature for the biases and errors that they elicit. It was also demonstrated that by applying a dual process framework, a way of conceptualising the heuristic and analytic processes that govern reasoning competence is provided. The experimental research presented in this chapter has two functions. Firstly, to examine the relationships between statistical and deductive reasoning, identifying any similarities or differences between them, and to investigate their relationships with cognitive ability. Secondly, to develop law of large numbers reasoning and selection task materials and ensure that they have the appropriate properties for the training studies that follow in Chapters 4 and 5.

This introduction will provide a brief overview of the different arguments proposed in relation to rational responding on a variety of critical reasoning tasks, particularly in terms of decontextualised reasoning and the different findings in the literature as to which individual differences variables predict this critical thinking skill. Under a dual process account of reasoning, decontextualised reasoning is an ability associated with System 2's analytical thinking strategies. By gaining further evidence of the factors which may facilitate this skill, more knowledge of the System 2 processes which govern reasoning competence will also be obtained. This will be followed by the rationale and predictions for Experiment 1.

3.1.1 Individual differences and decontextualised reasoning

Traditionally, research in reasoning and decision-making has investigated performance using between groups designs to examine the different effects of various experimental manipulations. More recently individual differences research has been conducted in order to identify factors which facilitate and inhibit reasoning performance on a variety of tasks (Stanovich & West, 1998a;1998b; Klaczynski, Gordon, & Fauth, 1997; Newstead, Handley, Harley, Wright, & Farrelly, 2004). This research has provided evidence in relation to the underlying processes which mediate reasoning competence in an individual, but at the same time has contributed to the development of modern dual process theories. For instance Stanovich and West (1998a; 1998b) have consistently found that only individuals of high ability are able to reason deductively on the arbitrary version of the selection task. Stanovich and West argued that it is these people that can override or inhibit the fundamental computational bias of System 1 which cue the heuristic (p/q) response. System 2, the analytic system is related to intelligence therefore it is the higher ability individuals that can decontextualise the elements of the problem from the scenario and reason deductively. Cognitive ability differences are attenuated on deontic problems as it is not necessary to decontextualise from the scenario to achieve the correct solution. Both systems cue the same correct response.

Newstead et al. agreed that very high ability participants were more able to solve the arbitrary selection task but they proposed a different explanation which incorporated a two-stage theory of decontextualisation. They suggested that the conditional rule must be abstracted from the scenario first which will lead to consistent responding, usually p/q. Secondly, they proposed that invited inferences such as the assumption that conditionals are biconditionals must be resisted in order to attain the correct response. Newstead et al. found that participants who attained the first level were still of higher ability when compared to inconsistent responders on the task.

Decontextualised reasoning has been identified as a key ability in critical thinking, necessary for optimum responding on most of the tasks discussed in this thesis. As discussed in Chapter 1, according to Stanovich (1999) the ability to decontextualise on a variety of reasoning tasks is linked which indicates that it is a domain-general dispositional trait related to the ability to reason independently of prior beliefs. Stanovich and West (1998b) conducted a series of studies to investigate the relationships between cognitive ability and thinking dispositions measures and performance on a variety of critical reasoning tasks. They employed tasks which they suggested involved separating prior knowledge or beliefs from the logic of the problem in order to achieve the normative response: syllogistic reasoning tasks, selection tasks, statistical reasoning and the Argument Evaluation Test (Stanovich & West, 1997). Under a dual process account these tasks induce conflict between the two systems of logic and belief.

The syllogistic reasoning tasks they used consisted of conclusions that contradicted world knowledge when the syllogism was valid and were consistent with world knowledge when the syllogism was invalid. In dual process terms there would then be conflict between the systems of logic and belief when attempting these problems. In order to achieve the correct response an individual would have to decontextualise their knowledge of the world from the logical format of the task. The selection tasks involved content which was real-life but arbitrary and as discussed in Chapter 1, correct solution rates on these tasks are usually less than 10%. The few people who solve this task correctly must resist the linguistic cues to test only the items named in the rule (Stanovich, 1999).

The statistical reasoning problems used by Stanovich and West (1998b) involved making an inductive inference in a simulation of a real-life decision. The information that was presented relevant to the decision was conflicting. One type was statistical and the other was a concrete case or personal experience that pointed in the opposite direction. Several of these statistical problems were adapted from Fong et al.'s (1986) studies. Judgement on these tasks has been found to be overly influenced by vivid but unrepresentational personal evidence and to be under influenced by more representative and diagnostic, but pallid, statistical evidence (Nisbett & Ross, 1980). According to Stanovich (1999) the participant must look beyond the vividness and immediacy of the testimonial evidence in order to realise that the abstract statistical information has greater validity.

Stanovich and West's (1997) Argument Evaluation Test was also discussed in Chapter 1. This task assesses an individual's ability to evaluate the quality of an argument independent of their feelings and personal biases about the proposition at issue. To perform well on this task, participants have to separate their prior opinion on the issue from the evaluation of the argument.

Stanovich and West (1998b) reported that all the tasks correlated with each other which indicated that those participants who respond normatively on one type of task tended to respond normatively on the other. In other words, departures from normative responding were not due to non-systematic performance errors but were due to systematic limitations in processing. This is important evidence for dual process theories of reasoning as it would indicate that analytic responding on all these tasks is being facilitated by the same system. However, there is conflicting evidence as to which individual difference factors facilitate this ability and to what extent (Newstead et al., 2004; Stanovich, 1999; Klaczynski, 1997).

Stanovich and West (1998b) found that cognitive ability and thinking dispositions were predictive of reasoning performance on all the tasks. Cognitive ability was the strongest unique predictor. However they showed that thinking dispositions were able to predict individual differences on the tasks even after cognitive ability was partialed out. They proposed that individuals with greater algorithmic capacity should show an enhanced ability to reason independently of prior belief. These individuals are also able to flexibly contextualise and decontextualise the problem dependent on what is necessary to resolve it. However thinking dispositions incorporate an individual's goals and beliefs therefore reasoning behaviour is also partly dependent on this intentional level of processing.

Conflicting findings were reported by Klaczynski, Gordon, and Fauth (1997). They agree that decontextualised reasoning is a crucial component of critical thinking but they proposed that cognitive ability was not associated with it. Klaczynski and Gordon (1996) reported that adolescents used sophisticated reasoning strategies only when such strategies led to conclusions which were consistent with their existing beliefs. Klaczynski, Gordon, and Fauth (1997) expanded on this and investigated the extent to which individual differences in cognitive ability and information-processing style were related to the capacity to reason using three critical thinking competencies: the law of large numbers, intuitive analyses of covariance, and dissection of hypothetical experiments for flaws.

Over a series of studies, participants were presented with critical reasoning problems designed to elicit the utilisation of the three competencies that were enhancing, threatening or neutral to each individual's occupational goal. Klaczynski, Gordon, and Fauth proposed that the cognitive mechanism responsible for reasoning biases involved the amount of information processing effort individuals exerted in processing goal-relevant information. Goal-enhancing information is readily assimilated to pre-existing belief systems and is processed at a relatively cursory level. Goal-threatening information on the other hand activates more sophisticated reasoning strategies in order to refute the evidence.

Klaczynski, Gordon, and Fauth found that general ability measures, such as verbal ability, were dissociated from measures of information processing style, and although they predicted the amount of statistical reasoning that an individual engaged in, they did not predict the degree to which individuals were biased by belief. Reasoning biases were associated with individual differences in rational processing as measured by Epstein et al.'s (Epstein, Pacini, & Heier, 1996) Rational-Experiential Inventory. They concluded that decontextualized reasoning was a function of an array of personal dispositions distinct from intelligence.

Stanovich and West suggested that one of the reasons that they found stronger relationships between cognitive ability and the tendency to evaluate evidence independent of prior beliefs, compared to Klaczynski et al. for example, is that the tasks they used resulted in different indices. Klaczynski et al.'s index is a direct measure of belief bias whereas Stanovich and West's is a measure of the ability to reason in situations in which prior beliefs may be interfering. They suggested that their measure combines context-free reasoning ability with the ability to ignore belief bias. By using the tasks from both bodies of literature in this study any relationships between them can be explored.

3.1.2 Aims of Experiment 1

This study has three main aims. Firstly to examine the relationships between the reasoning and decision making tasks. By employing a variety of tasks, we can explore Stanovich's proposal that normative responding on different problems is related. Secondly, we can also investigate the relationships between performance on the tasks and measures of cognitive ability and thinking dispositions in an attempt to test the different claims made by Stanovich and Klaczynski. Stanovich claims that ability predicts decontextualised reasoning whereas Klaczynski argues that it is a skill predicted by thinking dispositions. By employing tasks and measures utilised in both bodies of research, direct comparisons can be made.

The third aim is to develop the law of large numbers and selection task materials for the following training experiments. In order to investigate the effects of training on reasoning. it is necessary to have an understanding of the normative response on the task participants are being trained on, the biased responses that typically occur and the underlying processes which facilitate or mediate the different types of response. The tasks chosen for this individual differences study are all documented in the literature for their elicitation of differential responding according to analytic and heuristic processes as discussed in Chapter 1. Gaining further knowledge of the processes underlying reasoning competence will assist in the development of effective critical thinking training.

Findings in relation to performance and errors on the following tasks have been interpreted under dual process accounts of reasoning. Klaczynski's everyday reasoning experiments involving manipulation of beliefs have elicited two different types of responding cued by logic and belief on LLN problems and experiment evaluation problems. Performance patterns on arbitrary and deontic selection tasks have been described under the two systems, supported by individual differences in cognitive ability and thinking dispositions. The belief bias effects found in syllogistic reasoning involving real world content and performance on Stanovich and West's Argument Evaluation Test have also provided more evidence for two processes of thinking as discussed in Chapter 1. This study aims to evaluate dual process theories in the context of both formal and everyday reasoning tasks which will then allow further investigation into the relationships between normative and biased responding on each of the tasks.

The research reported in Chapters 4 and 5 examines training in two domains: statistical or law of large numbers reasoning, and selection task reasoning. Individuals generally invoke the LLN principle only under supportive conditions such as when cues are present in the problem (Fong et al., 1986). Klaczynski's (Klaczynski, 1997; Klacyznski, Gordon, &

Fauth, 1997) studies have consistently found that participants are more likely to use the LLN principle when evidence within the problem is threatening or inconsistent with their goals or beliefs. Klaczynski also found that ability predicted the amount of statistical reasoning an individual engaged in but thinking styles predicted the amount of bias independently. However Stanovich and West argue that cognitive ability and thinking dispositions both predict performance on the task. In Experiment 1 we replicate Klaczynski's studies and investigate the extent to which these individual differences variables influence differential responding on these everyday reasoning problems in order to understand how training may impact on reasoning performance.

The second group of training studies will focus on reasoning on selection task problems. The low solution rates on tasks involving arbitrary content has been consistently found for the last thirty years. The facilitation effects of deontic content have also been much explored and have led to context-dependent theories of reasoning such as pragmatic reasoning schemas (Cheng & Holyoak, 1985) and social contract theory (Cosmides, 1989). Again it is necessary to know what factors facilitate optimum responding on these tasks and the individual difference variables that may mediate an individual's performance, in order to understand how different training procedures may impact.

A second issue is that Klaczynski and Stanovich, whilst both finding a relationship between thinking styles and reasoning, employed different measures of disposition in their studies. Both thinking dispositions measures will be tested. Firstly the Thinking Dispositions Questionnaire (Stanovich & West, 1997; 1998b) which was found to be a unique predictor of the ability to evaluate arguments in the face of prior belief. High scores on a thinking disposition composite score derived from the scale reflects the tendency for more flexible, open-minded thinking styles. Associations were found between scores on the TDQ and cognitive ability indicating that the two constructs are linked. However Newstead et al. (2004) failed to find any such associations.

Secondly, Epstein, Pacini, Denes-Raj, & Heier (1996) provided evidence for the reliability and validity of a self-report measure of individual differences in intuitive-experiential and analytical-rational thinking, the Rational-Experiential Inventory which has subsequently been used as a measure in studies of bias in statistical reasoning and argumentation. Klaczynski, Gordon and Fauth (1997) reported that more rational individuals as measured by the REI were less biased in their reasoning style. It must be noted that rationality as measured by the REI corresponds to Evans and Over's (1996; Evans, Over, & Manktelow, 1993) Rationality 2, not total rationality. Klaczynski et al. found no relationship between responses on the REI and ability. However Newstead et al. found a correlation between rationality and ability but this was not replicated in any of their other studies. Evidence for the reliability of the shortened version of the Rational-Experiential Inventory (Pacini & Epstein, 1999) using a British population has been provided by Handley, Newstead, and Wright (2000) who also demonstrated a dissociation between information processing style and measures of general intelligence. By using both measures in a within subjects design it will clarify some of the differences evident in the previous research, such as the relationships with cognitive ability and reasoning tasks.

3.1.3 Predictions

There are several hypotheses to be examined. This study is an extension of Stanovich and West's (1998b) who found that normative responding on a variety of critical reasoning tasks was linked. Similar tasks will be used but in addition everyday reasoning problems designed to evoke belief-based responses will be included. If normative responding on all the tasks is related then this should extend to these tasks also. Bias on the everyday reasoning problems can also be computed therefore it will be possible to investigate whether bias on these tasks is related to bias in syllogistic reasoning, selection task performance and on the AET. If as Stanovich and West suggested the departure from normative responding is due to systematic limitations in processing and not random performance errors then these biases should also be related.

Individual differences will be explored in relation to performance on the critical reasoning tasks. According to Stanovich and West (1997; 1998b) both cognitive ability and thinking dispositions predict reasoning performance, with ability accounting for the larger proportion of the variance followed by thinking style. Taking this viewpoint then both measures will correlate with performance on the reasoning tasks in this study. However cognitive ability will have the strongest relationships. On the other hand, Klaczynski (1997; Klaczynski, Gordon, & Fauth, 1997) found that ability and thinking styles were not associated with each other. They found that biases in reasoning were predicted by information-processing style and ability measures predicted only the amount of statistical reasoning the individuals engaged in. In accordance with this view, ability and thinking styles measures will not be associated with each other. Ability measures will be associated with reasoning performance and thinking styles measures will be associated with bias, more rational individuals being able to inhibit biased responses. Stanovich and West suggested that one of the reasons that they find stronger relationships between cognitive ability and reasoning independently of prior belief than Klaczynski is that the different reasoning tasks used were measuring different effects. By using both Klaczynski's and Stanovich's tasks the relationships between the two tasks may be examined.

Everyday reasoning tasks designed to evoke LLN reasoning will be presented to the participants in this study. The aim is to replicate the findings of Klaczynski and his associates which illustrate that when presented with arguments that are inconsistent with an individual's goals or beliefs, the individual will use more sophisticated reasoning strategies to refute the evidence. However when the evidence is consistent with prior beliefs or goals then the reasoning utilised is cursory. Experiment Evaluation problems will also be used in which participants have to detect the flaws in hypothetical experiments. Again the problems will be designed to elicit belief-motivated responses.

On selection task performance, high ability scores will be related to correct responding on the arbitrary task. There will be no ability differences between solvers and non-solvers on the deontic tasks as according to Stanovich and West (1998a; 1998b; 1999) the correct answer is pragmatically and analytically cued. However, any association between ability and performance on the deontic tasks may also be due to analytic responding (Newstead et al., 2004).

3.2 Method

Design

This is a correlational study designed to investigate the relationships between information processing styles as measured by Pacini and Epstein's (1999) 40-item version of the Rational-Experiential Inventory and Stanovich and West's (1998b) Thinking Disposition Questionnaire; cognitive ability, statistical and experiment evaluation reasoning (Klaczynski et al.), Stanovich's Argument Evaluation Test, syllogistic reasoning and selection task performance.

Participants

A total of fifty participants took part in the study, thirty-eight female (mean age 20.95) and twelve male (mean age 22.17). Average age of the total sample was 21.24 (St. Dev. = 3.06). They were all undergraduates from the University of Plymouth, studying a variety of subjects taking part for course credit or cash payment.

Materials

Rational-Experiential Inventory

The rational part of the scale was originally based on the Need for Cognition scale which was developed by Cacioppo and Petty (1982). This is described by them as 'the tendency to for an individual to engage in and enjoy thinking'. The experiential scale was constructed to reflect the intuitive-processing counterpart of the Need for Cognition scale and was developed by Epstein, Pacini, Denes-Raj, and Heier (1996) and named the Faith in Intuition scale. Epstein et al. have reported that the factor structure is robust and the two scales correlate only minimally with each other (.07) which supports the claim that the two processing styles are independent of each other and operate in parallel.

More recently, the inventory has been redesigned and shortened to a 40-item questionnaire (Pacini & Epstein, 1999) with 20-items on the Rational scale and 20-items on the Experiential scale. This updated version of the Rational-Experiential Inventory has been reported to have a robust factor structure, very reliable subscales (.73 - .90, Handley, Newstead, & Wright, 2000) and highly significant test-retest correlations of .60 - .88 (Handley et al.). Within each scale there are two subscales, ability and engagement. Rational Ability refers to the confidence and ability with which one carries out a logical task, for example 'I have no problems thinking things through carefully'. Rational Engagement refers to the reliance and enjoyment one has in doing a logical thinking task, for example 'I enjoy thinking in abstract terms'. On the other hand, Experiential Ability refers to the confidence and ability one has in using their intuition, for instance 'When it comes to trusting people, I can usually rely on my gut feelings', whilst Experiential Engagement refers to the reliance on and enjoyment one has in using their intuition, for example 'I like to rely on my intuitions'. There are ten items per subscale (see Appendix A1.1 for the full scale).

The response format for all the items is on a five-point rating scale that ranges from 'definitely not true of myself' to 'definitely true of myself'. Half of the items are positive in direction and the other half are negative. Negative items are reverse scored. Overall Rationality and Experientiality scores are obtained by summing the participants' scores on the corresponding ability and engagement scales.

Thinking Dispositions Questionnaire

Participants were required to complete a 30-item questionnaire used by Stanovich and West (1998b, Experiment 1) consisting of items from a number of subscales as follows. The response format was on a 5-point scale (1=Strongly Disagree, 5=Strongly Agree). See Appendix A1.2 for the full questionnaire.

Actively Open-Minded Thinking Scale – consists of ten items that tap the disposition toward reflectivity, willingness to consider evidence contradictory to beliefs, and tolerance for ambiguity combined with a willingness to postpone closure. Design of the items was influenced by a variety of sources in the critical thinking literature but mostly from the work of Baron (1985, 1988), who has emphasized the concept of actively open-minded thinking through the cultivation of reflectiveness rather than impulsivity, the seeking and processing of information that disconfirms one's belief, and the willingness to change one's beliefs in the face of contradictory evidence. Examples of some of the items include: 'If I think longer about a problem I will be more likely to solve it.' and 'People should always take into consideration evidence that goes against their beliefs'.

Counterfactual Thinking Scale – consists of only two items devised by Stanovich and West (1998b) designed to tap counterfactual thinking. They were; 'My beliefs would not have been very different if I had been raised by a different set of parents' and 'Even if my environment (family, neighbourhood, schools) had been different, I probably would have the same religious views.' Both items were reverse scored so that higher scores indicate counterfactual thinking.

Absolutism – Stanovich and West adapted this scale from the Scale of Intellectual Development (SID) developed by Erwin (1981, 1983). Nine items were chosen which were designed to tap into the early stages of intellectual development and which are characterized by an absolutist orientation e.g. 'It is better to simply believe in a religion than to be confused by doubts about it' and 'right and wrong never change'.

Dogmatism – this subscale consisted of three items such as: 'Of all the different philosophies which exist in the world there is probably only one which is correct'.

Paranormal Beliefs – there were six items on the paranormal beliefs subscale, two concerned with belief in astrology and four concerned the belief in the concept of luck. For example: 'It is advisable to consult your horoscope daily' and 'The number 13 is unlucky'.

Thinking Dispositions Composite Score – A thinking dispositions composite score (TDC) was formed by summing the scores on the AOT and Counterfactual Thinking scales and then subtracting the sum of the scores on the Absolutism, Dogmatism, and Paranormal scales. Thus high scores on the TDC indicate open-mindedness, cognitive flexibility and a sceptical attitude. Low scores indicate cognitive rigidity and lack of scepticism (Stanovich & West, 1998b).

Cognitive Ability

In previous research, SAT scores have been used as a measure but as these are not available for the student population in the United Kingdom, the AH4-Group test of General Intelligence was administered (Heim, 1967). This test is designed for use with a cross section of the adult population. Test-retest reliability has been reported at 0.919 with retesting after one month. The test consists of two parts, each containing 65 items. Part one is composed of both verbal and numerical problems, part two is comprised of problems in diagrammatic form. Correlations between scores on Part One and scores on Part Two have been reported to range between 0.60 to 0.81 (Heim, 1967) and in this study the correlation was 0.56 ($p < .001$, two-tailed). Each section is timed and must be completed in 10 minutes.

Argument Evaluation Test (Stanovich & West, 1997)

The AET is a two-part measure including 23 items (see Appendix A1.3 for the full AET). For this study the items were modified for a British population and further items were added so the measure consisted of 25 items. The first part contains 25 propositions relating to real social and political issues of which people hold varying and possibly strong beliefs. Participants indicate their degree of agreement in each statement e.g. 'It is more dangerous to travel by air than by car' in order to ascertain participants' prior beliefs on each issue.

The second part consists of a set of instructions which introduce the participants to a fictitious individual called 'James' whose arguments they have to evaluate. There are 25 items which correspond to the propositions in the first part of the AET. Each argument starts with James stating a belief about an issue followed by a justification for his belief. A critic then presents an argument to counter this justification which the participants are informed to assume is correct. Finally James makes another rebuttal to the critic's argument which the participants are also informed to assume is correct. The participants are asked to evaluate the strength of James's rebuttal to the critic's argument. They are reminded to focus on the quality of James's rebuttal and to ignore whether or not they agreed or disagreed with James's original belief. Participants are asked to rate on a scale of 1-5 (very weak, weak, neither weak or strong, strong and very strong respectively) the strength of the rebuttal.

Participants' performance on the Argument Evaluation Test is calculated against an objective measure of argument quality. Stanovich and West (1997) used eight expert raters to evaluate the twenty-three items on their test and their median score for each item was used as the objective score to be compared against. They achieved a median correlation between the judgements of 0.74. In other words, the scores reflected the strength of the argument rebuttal without prior beliefs affecting any judgement. But as this measure had been anglicised, one of the original items had been removed and three new items had been added, a new objective index was required.

Four full-time members of staff at the University of Plymouth, all highly qualified in the field of human reasoning and decision making, one philosopher and the principal researcher rated the arguments. The median correlation between the six experts was 0.65 (see Appendix A1.4 for full correlation table). The medians of the scores for each item were compared to the medians used by Stanovich and West on the twenty-two items that remained and were found to correlate at 0.78. Therefore the revised median scores served as the objective index of argument quality (argument quality variable) and were then used for the regression analyses that follow.

The median scores for the participants on the twenty-five items ranged from 1 to 5. Six rebuttals received a median score of 1, four a median score of 1.5, four a score of 2, four a score of 4, three a score of 4.5 and four received a median rebuttal score of 5. A correlation between the sample's mean prior belief scores and the objective argument quality index was not significant at .37 ($p<.05$, one-tailed). The absence of a correlation shows that the items have been constructed so that objectively strong and weak arguments are equally associated with the prior beliefs that participants are thought to endorse. The correlation between the sample's mean prior belief scores and mean rebuttal scores for each item was significant at .68 ($p<0.001$), which illustrates that people are influenced by their beliefs when judging the rebuttal quality. The correlation between mean rebuttal scores and objective argument quality was significant at .80, which indicates that the participants were also greatly influenced by objectivity.

Separate regression analyses were run on each participant's responses to observe individual differences in participants' relative reliance on objective argument quality and prior belief. In other words, a separate multiple regression equation was constructed for each participant. The participant's evaluations of argument/rebuttal quality served as the criterion variable in each of the 50 separate regression analyses. The 25 evaluation scores were regressed simultaneously on both the 25 argument quality scores and the 25 prior belief scores. The regressions resulted in two beta weights for each participant, one for objective argument quality and one for prior belief. The former beta weight is the primary indicator of the ability to evaluate arguments independent of one's beliefs.

The beta weights for objective argument quality and prior beliefs correlated negatively at -.30 ($p<.05$) which would indicate two separate predictors of argument quality. The mean beta weight for prior beliefs as a predictor was .273 (SD=.209) with a minimum beta weight of -.29 and a maximum of .645. Sixteen out of the fifty beta weights achieved significance (Stanovich and West reported a mean of .151, st.dev. .218). The mean was significantly different from zero ($p<.05$, one-tailed). Only five out of the fifty participants had beta weights less than zero.

The mean beta weight for objective argument quality as a predictor was .370 (SD=.229) with a minimum beta weight of -.12 and a maximum of .764. Twenty-six of the beta weights achieved significance (Stanovich and West reported a mean of .330, st.dev. .222, 349 participants). Again the mean was significantly different from zero ($p<.01$, one-tailed). Only three out of fifty participants had beta weights less than zero.

On inspection of the data there are individual differences in the reliance of prior beliefs and objective argument quality when evaluating arguments. The next stage of the analysis was to split the participants in half using the beta weight scores for objective argument quality as this is viewed as the primary measure of the participant's ability to reason independently of their own beliefs. A median split of the sample resulted in 25 participants with a mean beta weight score of .561 for objective argument quality (termed HIARG because of their high reliance on objective argument quality) and 25 participants with a low mean beta weight score of .178 for objective argument quality (termed LOARG because of their low reliance on objective argument quality). The difference in beta weights was highly significant, $t(24)=-10.62$, $p<0.001$.

The mean beta weight for prior belief in the HIARG group was .213 which was significantly lower than the mean beta weight for the LOARG group at .334, $t(24)=2.25$, $p<.05$. This indicates that the median split reliably partitioned the sample into a group of participants who relied more on objective argument quality for argument evaluation decisions (HIARG) and a group who relied more on their prior beliefs (LOARG).

Statistical Reasoning

The problems used to assess participants' understanding of law of large numbers were adapted from Klaczynski, Gordon, and Fauth (1997; see also Klaczynski & Fauth, 1997; Fong, Krantz, & Nisbett, 1986). In part one of this study, participants were asked to indicate their occupational goal. In part two, participants were presented with nine problems designed to elicit law of large numbers reasoning. These consisted of three different problem types: goal-threatening problems involved arguments which were threatening to a participant's occupational goal, goal-enhancing problems involved arguments which were complimentary, and goal-neutral problems involved arguments related to a different occupational goal. See Appendix A1.5 for the instructions and the nine LLN problems.

Within each problem type there were three structures of problem. The first type of problem involved a hypothetical argument in which an actor makes a generalisation from a single instance or from a very small sample of observations. The second type involved arguments based on both large and small samples. One actor draws a conclusion based on a small sample and personal experience and the second actor draws a contradictory conclusion from a larger sample of evidence. The third type of problem involved a hypothetical actor drawing a hasty conclusion from a small sample. Table 3.1 presents an example of each structure of problem involving arguments which are either neutral, enhancing or threatening to a person's occupational goal.

Two forms of the problem were constructed where evidence that was goal-threatening in one form became goal-enhancing in the other, and vice versa for the second form. Participants were presented with the same problems except for the occupation involved and the problem form. The nine problems were presented in a random order for each participant, the only constraint being that no two problems of the same type were presented consecutively.

Participants indicated on a 9-point scale how strong they thought the conclusion was based on the evidence used (1=extremely weak, 9=extremely strong). They were also asked to indicate on a 9-point scale how convinced they were by the argument (1=extremely unconvinced, 9=extremely convinced). Total scores on each rating scale were calculated separately for goal-enhancing, goal-threatening and neutral problems. Scores on each rating scale could range from 3 to 27, for each problem type. Klaczynski et al. referred to these as the 'evidence evaluation' and 'persuasiveness' ratings, respectively.

Structure 1: Goal-neutral

Amy is thinking about being a chemist, and her friend Jill is trying to talk her into it. Jill's argument is, "Look, my father was a chemist, and after 20 years of being a chemist, he's still energetic, trying to do his best, and is very happy. It's a great job to do. You can just take one look at my dad and you'll know that he's got a great job."

Structure 2: Goal-enhancing

Lisa and Roy are having an argument over whether exploration geologists are more likely to get divorced than other people. Roy claims the marriages of exploration geologists are much less satisfying than those of other people and frequently end in divorce, but Lisa believes that exploration geologists' marriages are very happy.

Roy:
I've been working as a lawyer for over 15 years, and handle people's divorces all the time. During all that time, I've seen hundreds of couples in bad marriages that wind up in divorce and they've been from just about every occupation you can imagine, but most of those divorces by far have involved exploration geologists, more than any other occupation. I've seen enough to know that a marriage is pretty likely be unhappy, unsatisfying, and to split up when an exploration geologist's involved.

Lisa:
You're wrong and you know it! Two of my sisters are exploration geologists, remember? I talk to them every day and they're always talking about how happy they are in their marriages. In fact, I know several people who are exploration geologists and they have the most satisfying marriages I've know of. I've seen it for myself, you haven't! If you've seen the exploration geologists I've seen, it's pretty easy to figure out that exploration geologists are a good bet to have life-long marriages.

Structure 3: Goal-threatening

John works for a London testing company. His company has been hired to measure the IQs or intelligence of Ministry of Defence employees. John himself has been asked to give an IQ test to 10 people who are MOD employees. Two people were secretaries, two were doctors, 2 were custodians, 2 were architects, and 2 were dentists. He uses the best IQ test available, one that is respected across the country. After he had tested each person once, he put his findings in a table:

Person:	IQ:
Secretaries	98 and 101
Custodians	90 and 97
Doctors	115 and 117
Architects	106 and 111
Dentists	102 and 108

Because a person with average intelligence should have an IQ of around 100 and because IQs below are low, especially for most professions, John concludes that custodians are much less intelligent than the other groups of people he tested.

Table 3.1 Examples of the three structures of argument involving either neutral, enhancing or threatening conclusions taken from Klaczynski, Gordon, and Fauth (1997)

The participants were then requested to write explanations of why they believed the arguments were weak or strong in no more than two or three sentences. These explanations were scored using a 3-point system developed by Fong et al. A score of zero was given if the response contained no indication of statistical reasoning; there was no mention of sample size, random variability, or the role of probability; for example, 'I believe Jill because she has seen it with her own eyes and hasn't just read it out of the newspapers'. A score of 1 was given when the participant referred to the law of large numbers, but vaguely, for example, 'It's good at concluding what Jill thinks about her dad, but is not weak or strong about chemists in general because there are no other chemists being described'. In this case the participant implied that he or she was using statistical reasoning, but was not explicit about the statistical basis of his or her reasoning. If a participant scored 2 for a response, it meant the LLN principle was clearly applied in the explanation. For example, 'Jill's argument is pretty bad because she used just one person. If you make that kind of generalization, it ought to be based on hundreds or maybe thousands of people'.

For each problem type, scores were collapsed and added together to produce three total reasoning scores that range from 0-6. A random sample of 30 out of the 50 participants' responses were scored by two independent judges and the researcher, and based on the responses of 30 participants (270 problems) total agreement was achieved for 84%. The principal researcher scored the remaining 20 participants' items. Items were collapsed across problem type (neutral, enhancing and threatening to a participant's occupational goal).

Bias scores - A bias score was computed for reasoning on the LLN problems by subtracting scores on the goal-enhancing problems from scores on the goal-threatening problems. Positive differences would indicate goal-biased reasoning. Biases in rating scores were then calculated by subtracting ratings on goal-threatening problems from those on the goal-enhancing problems. Again the larger the difference, the greater the bias.

Experiment Evaluation Task

Nine problems were constructed based on those used by Klaczynski, Gordon and Fauth (1997). As for the LLN problems, there were three types of problem, one of each was goal-enhancing, one goal-neutral and one goal-threatening. Two forms of each problem were created so a problem that was goal-enhancing in one form was goal-threatening in the other.

The scenario for each problem consisted of a brief description of the hypothetical individuals who participated, their occupations (including the participant's occupation), the methods used to conduct the research, and the findings and conclusions drawn from the research. Within each problem there were three structures which differed in the threat to internal validity built into the scenario. See Appendix A1.6 for the instructions and nine problems.

Structure one problems involved descriptions of psuedo-experimental research in which the primary independent variable was confounded with another variable. Structure two problems described research comparing the participant's occupation with other occupations on some variable of interest, however a selection confound was built into the research, such that the occupational groups were selected from very different populations. In Structure three problems, the construct validity of the dependent variable was suspect. Table 3.2 presents an example of each structure of experiment within each problem type .

Participants were informed prior to the presentation of the experiment evaluation problems that they would be presented with several summaries of actual psychological and sociological research. They were told that the research involved participants from several dozen occupations, one of which was their intended occupation. It was further explained that many of the findings that pertained to other occupations had been deleted from the summaries and that only information relevant to the participant's occupation and a few other, randomly selected, occupations would be presented.

Structure 1: Neutral

Professor Y. wanted to find out whether individuals who worked in any one occupation were more or less socially skilled than individuals who worked in another. She randomly selected 40 dentists, 40 doctors and 40 teachers from all over the United Kingdom and invited them to a conference and dinner together. The teachers couldn't make it on the same day therefore their conference/dinner was held two weeks later. At the conferences and dinners, individuals were watched by trained researchers to observe their behaviours. For example, introducing themselves and introducing others. On collation of all the data, Professor Y. found that dentists showed the most social behaviours, closely followed by doctors, but with significantly less social behaviours were the teachers. In reporting her findings, Professor Y. concluded that dentists are more socially skilled individuals than doctors or teachers.

Structure 2: Enhancing

Two well-known researchers have conducted several studies on the relationship between one's occupation and the tendency to conform to authority. The occupations of being a teacher and being an accountant were compared in one study. In this study, 50 teachers were contacted locally and agreed to be in the study. At the time of the study, accountants were hard to locate and contact. Thus, the researchers recruited 50 accountants from a conference on leadership that was taking place in a nearby city. Next, each person was brought to the researchers' laboratory. At the laboratory, both the accountants and the teachers were given several orders to complete menial tasks. For example, they were ordered to grade the papers of several dozen undergraduates. At the completion of the experiment, the researchers found that the teachers were far more likely to obey the commands than were the accountants. The accountants were much more likely to question the commands and to refuse. The researchers concluded that the experience of being an accountant results in a greater sense of independence and of inner strength than does being a teacher.

Structure 3: Threatening

A leading psychologist from a University Hospital conducted a study to look at overall satisfaction with life in individuals from different occupations. For the research she compared teachers, nurses, secretaries, accountants and waiters/waitresses. Sixty people were recruited from each occupation and for a fortnight they were each required to note down in a diary each positive thought that they had, day or night. At the end of the study period the results were collated and it was found that overall teachers had an average of 7.1 positive thoughts a day, nurses and waiters/waitresses had an average of 6.0 positive thoughts a day, secretaries had an average of 8.2 positive thoughts a day and accountants had an average of only 2.0 positive thoughts a day. From these results the psychologist noted that accountants had significantly less positive thoughts each day than teachers, nurses, secretaries or waiters/waitresses. She concluded that this indicated that accountants were overall much more dissatisfied with their lives than individuals from any of the other occupations.

Table 3.2 Examples of each structure of problem involving either neutral, enhancing or threatening conclusions taken from Klaczynski, Gordon, and Fauth (1997)

Following each problem, participants were asked to indicate on a 9-point scale how strong the researcher's conclusion was (1=extremely weak, 9=extremely strong), and how valid the experiment was (1=extremely invalid, 9=extremely valid). Total scores on each rating scale were calculated separately for goal-enhancing, goal-threatening and neutral problems. Scores on each rating scale could range from 3 to 27, for each problem type. These will be referred to the as the 'conclusion strength' and 'experiment validity' ratings, respectively.

Following this participants were asked to write an explanation of why they thought the research was weak/strong and valid/invalid in no more than two or three sentences. As with the LLN problems, participants received the same problems apart from form and occupational goals. No two problems of the same type were presented consecutively.

The scoring system for the experiment evaluation problems was the same as used by Klaczynski, Gordon, and Fauth and was again based on a 3 point scale. A score of 0 was given if the participant gave no indication that any threat to the validity of the experiment existed. That is the participant believes that the experiment is validly conducted or rejects the experimental evidence by simply asserting that it is true or false. For example, 'This is good research. The researchers had a good hypothesis and the research shows that they were right'. A participant scored 1 if he or she indicated their awareness of the confound built into an experiment but did not indicate that the existence of this confound made it impossible to make a straightforward interpretation of the findings. For example, 'This might not be true because they attended the conference on different days'. Finally a score of 2 was awarded when a participant indicated that the experimental confound made it impossible to interpret the findings of and draw conclusions from the research. For example, 'the experiment is invalid as the different occupations do not attend the conference on the same day. It therefore does not test what it set out to test'. Responses for a subset of 30 participants (270 problems) were coded by two independent coders and the principal researcher and the total agreement achieved was 87%. The principal researcher coded the items for the remaining 20 participants.

Bias scores - A bias score was computed for reasoning on the experiment evaluation problems by subtracting scores on the goal-consistent problems from scores on the goal-inconsistent problems. Positive differences would indicate goal-biased reasoning. Biases in rating scores were then calculated by subtracting ratings on goal-threatening problems from those on the goal-enhancing problems. In accord with the LLN ratings, the larger the difference the greater the bias.

Belief Bias Syllogisms

Two forms of syllogism were used for this task:

Form 1: No A are B,
Some C are B,
Therefore,
Some C are not A.

Form 2: Some A are B,
No C are B,
Therefore,
Some A are not C.

These two forms have been widely used in studies of belief bias (e.g. Evans, Barston & Pollard, 1983). Within each form, eight syllogisms with realistic content were employed, four with Empirical content and four with Definitional content. An example of each is shown in Table 3.3 (see Appendix A1.7 for full set).

Empirical	Definitional
No police dogs are vicious,	Some mice are timid.
Some highly trained dogs are vicious,	No tigers are timid,
Therefore,	Therefore,
Some highly trained dogs are not police dogs.	Some mice are not tigers.

Table 3.3 Examples of syllogisms involving Empirical and Definitional content

The syllogisms within each form and content led to 4 Valid Believable, 4 Valid Unbelievable, 4 Invalid Believable and 4 Invalid Unbelievable conclusions therefore making sixteen items in all.

Participants were given unlimited time to complete the syllogisms. The instructions were taken from Evans, Newstead, Allen, and Pollard (1994; Experiment 1) with the augmented instructions omitted.

Two separate indices were computed for logical responding and belief-based responding. Logic was computed by subtracting performance on the invalid syllogisms from the valid syllogisms and Belief was computed by subtracting performance on unbelievable problems from the believable ones.

Selection Tasks

Four selection task problems were completed by participants, two involving arbitrary rules and two consisting of deontic content. The first one was the abstract version of the Wason selection task (Wason, 1966) using letters and numbers (A, K, 8 and 5), and the second an arbitrary version of the Destination problem ('If Glasgow is on one side of the ticket, then train is on the other side of the ticket'). The version of instructions 'test whether the rule is true or false' was used as this doesn't appear to facilitate performance as other versions have been shown to (Stanovich & West, 1998a). See Appendix A1.8 for all four versions.

The two deontic versions were an anglicised version of the Sears problem in which Sears becomes Debenhams (any sale over £30 must be approved by the manager, Mr Jones') and an elaborate version of the Drinking-age problem with full scenario (Klaczynski & Laipple, 1993). For these tasks, 'violation' instructions were used to optimise facilitation. All the reasoning problems were adapted from Stanovich and West (1998a) and all were accompanied by a graphic choice. The order of the four alternatives that represented the choices p, not-p, q and not-q was different for each problem.

Procedure

Participants were tested in small groups of four to twelve. Testing took place in two sessions of one hour, approximately one week apart. This was to negate any effects of fatigue. In addition the nature of the study required participants to indicate beliefs and attitudes which may cue responses to the everyday reasoning problems and Stanovich's Argument Evaluation Test. In the first testing session, participants first completed the AH4-Group test of intelligence. As this was the only timed component of the study, participants were then free to complete the remaining tasks in session one at their own pace but in the order presented (see Appendix 3A-3H for materials). First they were asked to complete the Rational Experiential Inventory followed by Stanovich's Thinking Disposition Questionnaire and the first part of the Argument Evaluation Test. Their future occupational goal was ascertained by a series of questions which were embedded in the middle of the questionnaires. Participants were then required to continue with the selection tasks which were presented in the order of Wason, Destination, Sears and then Drinking Age. Finally, the sixteen syllogisms were presented in random order. Demographic information on age, gender and qualifications attained were also obtained.

In the second testing session, participants completed the law of large numbers problems followed by the Argument Evaluation Test and then the experiment evaluation problems. At the end of this session participants were thanked and debriefed.

3.3 Results

There were three main aims of this study which will be investigated in the following order: firstly, Stanovich's proposal that normative responding on the reasoning tasks is related. Logical performance on each of the reasoning and decision making tasks will be compared. Secondly, we wanted to investigate the relationships between individual differences measures of thinking dispositions and cognitive ability with performance on the reasoning tasks. Finally we needed to validate the materials for the training studies which follow in Chapters 4 and 5.

Table 3.4 presents the descriptive statistics for performance on each of the reasoning tasks. In syllogistic reasoning, participants achieved higher scores on problems involving no conflict between logic and belief, that is valid-believable and invalid-unbelievable conclusions than problems involving conflict, that is valid-unbelievable and invalid-believable conclusions. Table 3.4 also displays the mean score for performance on the law of large numbers reasoning problems. A bias score was computed by taking the LLN score on problems which were goal-enhancing from the score on problems which were goal-threatening as illustrated. The mean score overall on the experiment evaluation problems is also shown. Again a bias score was computed by subtracting the scores on the goal-enhancing problems from the scores on the goal-threatening problems. Finally the mean scores on Stanovich's Argument Evaluation Test are shown, more specifically, the mean indices for prior belief and for objective argument quality (see Method section for details of how the indices were calculated).

The selection tasks were scored by allocating a score of one for a correct card selection and minus one for an incorrect card selection. In other words, for p and not-q responses a score of one was given for each, and for not-p and q responses a score of –1 was given. For example, if the answer p/not-q was given then the total score was 2 for that item. Tasks involving deontic content yielded higher indices than ones consisting of arbitrary content as expected.

Tasks	Indices	Mean	Minimum	Maximum	Std. Dev.
Syllogistic reasoning	Conflict	4.52	1.00	8.00	1.53
	No conflict	5.68	1.00	8.00	1.65
LLN reasoning	LLN	6.40	0.00	13.00	2.61
	LLN bias	0.20	-4.00	3.00	1.54
Experiment evaluation reasoning	Exp. Evaluation	7.76	0.00	18.00	3.71
	EE bias	0.30	-4.00	4.00	1.84
Argument Evaluation Test	Prior Belief	0.27	-0.29	0.64	0.21
	Objective argument	0.37	-0.12	0.76	0.23
Selection Tasks	Arbitrary	1.42	-2.00	4.00	1.71
	Deontic	2.90	0.00	4.00	1.43

Table 3.4 Descriptive statistics for performance on the reasoning tasks

3.3.1 Normative responding on the tasks

Stanovich has shown that performance on a variety of critical thinking tasks is linked. According to Stanovich and West (1998b) individuals who give the normative response on one task tend to give the normative response on another. Note that for the following analyses only correlations are reported as the power of the study was too weak, due to the low number of participants, for more sophisticated investigation such as Factor Analysis. The first correlational analysis examined is illustrated in Table 3.5 which presents the correlations between performance on all of the tasks. Performance on the arbitrary versions of the selection task correlated with performance on the deontic versions and evidence evaluation reasoning (.44, $p<.01$ and .33, $p<.05$ two-tailed respectively). Performance on the syllogistic reasoning problems involving conflict between logic and belief was related to law of large numbers reasoning (.36, p<.05 two-tailed), experiment evaluation reasoning (.37, p<.05 two-tailed) and the objective argument quality index on the argument evaluation task (.30, p<.05 two-tailed). Scores on the objective argument quality index were also associated with experiment evaluation reasoning (.44, p<.01 two-tailed) and performance on the deontic problems (.27, p<.1 two-tailed). In other words, participants who were more objective on Stanovich's argument evaluation test were better at performing on syllogistic reasoning problems where logic and belief were in conflict, were better at finding the design flaws in hypothetical experiments, and were better at deontic reasoning tasks. There was an association between performance on the deontic selection tasks and the non conflict problems in syllogistic reasoning (.31, p<.05 two-tailed) which implies that the correct responses on both these tasks may be cued pragmatically as well as analytically. Surprisingly, logical performance on the deontic tasks is strongly linked to normative responding on the other tasks. Whist the correlations are positive the correlations are weaker for the arbitrary selection task.

	Selection tasks		Syllogistic reasoning		Everyday reasoning		AET
	Arbitrary	**Deontic**	**Conflict**	**No conflict**	**LLN**	**EE**	**Objectivity**
Arbitrary							
Deontic	0.44***						
Conflict	0.21	0.33**					
No conflict	0.23	0.31**	0.23				
LLN	0.11	0.38***	0.36**	0.15			
Experiment Evaluation	0.33**	0.31**	0.37**	0.05	0.15		
Objectivity	0.17	0.27*	0.30**	-0.08	0.23	0.44***	

*p<.1, **p<.05, ***p<.01 two-tailed

Table 3.5 Correlations between performance on each of the reasoning tasks (N=50)

It was expected that belief based responses on the tasks, i.e. the evidence evaluation task, LLN task, belief indices in syllogistic reasoning, heuristic responding on the arbitrary selection tasks and Stanovich's measure of prior belief on the argument evaluation task. would have been related with each other but no correlations attained significance (see Table 3.6 for correlations).

	Belief	Lln bias	EE Bias	Prior belief	Arbitrary (p/q)
Belief					
LLN bias	0.10				
Experiment evaluation bias	-0.25	-0.12			
Prior belief indices	0.20	0.09	-0.13		
Arbitrary (p/q)	-0.02	0.05	0.10	0.20	

Table 3.6 Correlations between biased responses on the tasks (N=50)

Table 3.7 presents the pattern of correlations between performance on problems involving conflict between logic and belief, and problems involving no conflict. Performance on law of large numbers reasoning and experiment evaluation tasks consisting of conclusions which are inconsistent with beliefs was related to tasks consisting of conclusions which are consistent with beliefs (.28, p<.05 and .39, p<.01 two-tailed for LLN and experiment evaluation respectively).

		Syllogistic		Law of Large Numbers		Experiment Evaluation	
		Conflict	**No conflict**	**Threatening**	**Enhancing**	**Threatening**	**Enhancing**
Syllogistic	**Conflict**						
	No conflict	0.23					
LLN	**Threatening**	0.20	0.15				
	Enhancing	0.30**	0.07	0.28**			
EE	**Threatening**	0.45****	0.14	0.10	0.17		
	Enhancing	0.11	0.13	0.18	0.07	0.39***	

*p<.1, **p<.05, ***p<.01, ****p<.001 (two-tailed)

Table 3.7 Correlations between problems involving conflict and no conflict

There is a strong association between performance on goal-threatening experiment evaluation problems and conflict problems in syllogistic reasoning (.45, $p<.001$, two-tailed). The relationship between conflict and no conflict problems in syllogistic reasoning just failed to reach significance.

3.3.2 Individual differences and normative responding

Another purpose of this study was to observe the relationships between thinking dispositions, cognitive ability and performance on the above reasoning tasks. Before investigating these relationships, there will be a brief overview of the psychometric findings in relation to the two thinking styles measures, the Rational-Experiential Inventory and Stanovich's Thinking Dispositions Questionnaire.

3.3.2.1 Rational-Experiential Inventory

The Cronbach alpha coefficient calculated for the REI was very high at .86. As may be seen in parentheses in Table 3.8 the internal reliabilities of the subscales of the REI are comparable with previous studies. See Pacini and Epstein (1999) who reported reliability coefficients of .90 and .87 for the Rationality and Experientiality scales respectively, and also high reliability coefficients for the four subscales, RA, RE, EA and EE (.77 to .81). Similar results were also reported by Handley, Newstead and Wright (2000). There was no correlation between the Rationality and Experientiality scales, which supports Epstein's (1994) claim that the two scales are independent of each other and indicative of two independent information processing styles.

	Rational Total	Rational Ability	Rational Engagement	Experiential Total	Experiential Ability	Experiential Engagement
Rational Total	(.87)	.87***	.86***	.02	.13	-.07
Rational Ability		(.84)	.50***	-.09	.07	-.20
Rational Engagement			(.80)	.13	.15	.09
Experiential Total				(.89)	.91***	.95***
Experiential Ability					(.74)	.73***
Experiential Engagement						(.87)

*p<.05 **p<.01 ***p<.001 (two-tailed)

Table 3.8 Intercorrelations and Reliabilities of the REI Scales. (N = 50)

3.3.2.2 Thinking Dispositions Questionnaire

The Cronbach alpha coefficient for the scale was only moderate at .53. Overall the REI has a higher internal consistency and appears a more reliable measure than the TDQ as found by Newstead et al. (2004). Table 3.9 displays the mean scores on each of the subscales. Also included is Stanovich and West's (1998) Thinking Dispositions Composite score which was calculated by subtracting the scores on three of the subscales (dogmatism, absolutism and paranormal belief) from the scores on the other subscales (counterfactual thinking and actively open-minded thinking).

	Mean	Minimum	Maximum	Std.Dev.
Actively open-minded thinking	41.90	33	49	4.14
Paranormal beliefs	11.00	6	24	5.18
Counterfactual thinking	7.02	2	10	2.21
Dogmatism	7.24	3	11	1.78
Absolutism	23.52	14	35	4.79
Thinking Disposition Composite	7.16	-18	27	11.90

Table 3.9 Means and standard deviations for subscales of the Thinking Disposition Questionnaire

3.3.2.3 Individual Differences and performance on the tasks

Of particular interest at this point are the associations between performance on all the tasks and cognitive ability and thinking dispositions. Table 3.10 presents the correlations between the two measures of thinking styles, cognitive ability and performance on the reasoning tasks.

	Scores	AH4 total	rationality	experientiality	tdc
Selection tasks	Arbitrary	0.27*	0.22	0.14	0.26*
	Deontic	0.25*	0.16	0.21	0.19
Syllogisms	Conflict	0.18	0.33**	-0.18	0.25*
	No conflict	0.20	0.18	-0.17	0.40***
Law of Large Numbers	LLN	-0.02	0.20	0.02	0.13
	Lln bias	0.04	0.08	0.18	0.03
	Evidence evaluation bias	-0.15	-0.14	0.21	-0.42***
	Persuasiveness bias	-0.16	-0.11	0.03	-0.36**
Experiment Evaluation	Experiment Evaluation	0.38***	0.02	-0.13	0.13
	E E bias	0.19	0.38***	-0.03	0.26*
	Conclusion strength bias	0.13	0.34**	-0.03	0.17
	Experiment Validity bias	0.18	0.42***	-0.00	0.24*
Argument Evaluation Test	Prior Belief	-0.06	0.02	0.19	-0.15
	Objective Argument Score	0.12	0.07	0.02	0.26*

*p<.1, **p<.05, ***p<.01 two-tailed

Table 3.10 Cognitive ability, thinking dispositions and performance on the reasoning tasks and bias (N=50)

Performance on both arbitrary and deontic versions of the selection task were marginally associated with cognitive ability (.27 arbitrary; .25 deontic, both p<.1 two-tailed). Performance on the arbitrary tasks was also related to the Thinking Disposition Composite score (.26, p<.1 two-tailed) as predicted by Stanovich. On the syllogistic reasoning tasks, performance on the conflict problems (valid-unbelievable, invalid-believable) was related to rationality as measured by the REI (.33, p<.05 two-tailed) and the TDC (.25, p<.1 two-tailed). In other words, participants who reported themselves to be more able and enjoy engaging in such tasks and were more open-minded, were better at correctly endorsing the problems where logic and belief are in conflict. Performance on problems which involved no conflict (valid-believable and invalid-unbelievable) was related to Stanovich's Thinking

Disposition Composite score (.40, p<.01 two-tailed). The correct endorsement of problems in which logic and belief are not in conflict was related to actively open-minded and flexible thinking styles. This finding is incongruent with Newstead et al.'s findings where they consistently found that the rationality scale correlated with nonconflict problems and only ability was associated with conflict items.

There were no associations found between law of large numbers reasoning and cognitive ability or thinking dispositions. Bias on the experiment evaluation rating scale was negatively associated with Stanovich's Thinking Disposition Composite score (-.42, p<.01 two-tailed). In other words, the more objective and flexible in the thinking style the less biased a person was when rating argument strength. A negative association between persuasiveness ratings and the Thinking Dispositions Composite score (-.36, p<.05 two-tailed) was also found. In other words, the more objective and flexible in your thinking style the less likely you are to find the arguments convincing. Klaczynski also found that thinking style predicted bias in ratings. Again no associations were found between bias and ability.

Performance on the experiment evaluation problems was found to be related to cognitive ability (.38, p<.01 two-tailed), and the computed bias score was related to rationality (.38, p<.01 two-tailed). Therefore the higher ability participants identified more problems in the hypothetical research, but participants who reported themselves to be of a more rational thinking style were actually more biased which is in direct contrast to what was predicted. The correlations presented in Table 3.10 also illustrate associations between the rationality scale and bias in ratings on both the conclusion strength and experiment validity scales (.34, p<.05 and .42, p<.01 two-tailed, respectively). These relationships with bias are also in the opposite direction to what was expected.

There was only one relationship found between the objective argument quality indices resulting from performance on Stanovich and West's Argument Evaluation Test, and the TDC. Participants who reported themselves to be more open-minded flexible thinkers relied more on objectivity when responding on the AET.

3.3.3 Materials validation

The second aim of this study was to develop the materials for the following training studies. We needed to replicate findings in the literature in relation to logical and non-logical response patterns in statistical, evidence evaluation, syllogistic and selection task reasoning.

3.3.3.1 Statistical Reasoning

The three-point system developed by Fong et al. (1986; Klaczynski & Gordon, 1996; Klaczynski & Fauth, 1997; Klaczynski, Gordon, & Fauth, 1997) was used to code LLN reasoning (see Method section for coding and reliability). Items were collapsed across problem type (goal-enhancing, goal-threatening and goal-neutral) and means and standard deviations are presented in Table 3.11. Reasoning items that included arguments threatening to a participant's goals evoked the most LLN reasoning as expected followed by items that involved goal-enhancing arguments. The use of LLN reasoning was lowest on the neutral items. A repeated measures Anova with problem type as the independent variable reached significance ($F(2, 98) = 3.03$, $MSE = 3.73$, $p=0.05$). An LSD post-hoc comparison showed that the only significant difference was between LLN reasoning on the neutral items compared to LLN reasoning on the threatening items ($p<.05$). There was no difference between goal-threatening and goal-enhancing problems. In other words, more sophisticated reasoning strategies are utilised on items that contain conclusions that are incongruent with one's beliefs as reported by Klaczynski, Gordon, and Fauth (1997), but only compared to neutral content.

	Mean	Minimum	Maximum	Std.dev.
Neutral	1.84	0	5	1.20
Goal-threatening	2.38	0	6	1.38
Goal-enhancing	2.18	0	5	1.17

Table 3.11. Means and Standard Deviations for LLN reasoning on the three problem types (N=50)

3.3.3.2 Experiment evaluation reasoning

The 3-point scoring system developed by Klaczynski, Gordon, and Fauth (1997) was used to score the Experiment Evaluation problems. Items were then collapsed across problem type (Goal-neutral, enhancing and threatening). Table 3.12 presents the means and standard deviations under each problem type. Mean scores are higher on problems involving goal-threatening conclusions as expected. That is participants appear to use more sophisticated reasoning strategies on these problems as in LLN reasoning problems. However, a repeated measures Anova with problem type as the independent variable failed to reach significance ($F(2, 98) = 2.24$, $MSE = 3.40$, $p > .05$).

	Mean	Minimum	Maximum	Std.Dev.
Neutral	2.34	0.00	6.00	1.48
Goal-threatening	2.86	0.00	6.00	1.84
Goal-enhancing	2.56	0.00	6.00	1.43

Table 3.12 Means and standard deviations for experiment evaluation reasoning across the 3 problem types (N=50).

It appears that we were not entirely successful at influencing the sophistication of reasoning strategies by manipulating occupational goals. There are a number of possible reasons for this that are considered in the general discussion.

3.3.3.3 Syllogistic reasoning

Table 3.13 illustrates the mean number of conclusions accepted (as a percentage) as a function of logic and belief. More items were endorsed for the valid items than invalid items; more believable items were endorsed than unbelievable items.

	Believable	Unbelievable
Valid	80	84
Invalid	71	38

Table 3.13 Percentage of Syllogistic conclusions endorsed as a function of belief and logic.

An 2 x 2 Anova (logic x belief) was performed which yielded a main effect of logic ($F(1, 49) = 39.01$, $MSE = 60.50$, $p<0.001$). That is participants endorsed more valid than invalid conclusions. The Anova also yielded a main effect of belief, that is they endorsed more believable than unbelievable conclusions ($F(1, 49) = 17.29$, $MSE = 16.82$, $p<0.001$). The interaction between the two variables was also significant ($F(1, 49) = 30.54$, $MSE = 25.92$, $p<.001$). In other words, when the conclusion is believable, participants are more likely to endorse it whether it is valid or not. When the conclusion is unbelievable, they will endorse it if it is valid but not when it is invalid.

A correlation was performed on the two indices of logic and belief (see Method section for how these indices are constructed) and it failed to reach significance (.08, $p>.05$) which, in support of Klaczynski, implies that the two are independent of each other.

3.3.3.4 Selection tasks

The pattern of correct responses corresponds to previous studies but with a high percentage of participants obtaining the correct response at 24% for the Wason selection task, 38% for the Destination problem, 56% for the Debenhams problem and 80% correct for the Drinking age task. The difference between performance on the deontic problems and performance on the non deontic problems was significant ($t(49) = -6.22$, $p<.001$) illustrating the large facilitation effects of deontic problems.

Table 3.14 presents the correlations between the selection tasks. The highest correlations are between the two abstract versions at .48 and the two deontic versions at .47 (both p<.001, two-tailed). Therefore the strongest correlations are between tasks in the same domain.

	Wason	Destination	Debenhams	Drinking age
Wason		.48***	.29*	.18
Destination			.42**	.39**
Debenhams				.47***
Drink				

*p<.05 **p<.01 ***p<.001 (two-tailed)

Table 3.14 Correlations between selection tasks (N=50)

3.4 Discussion

The main aims of this study were to investigate Stanovich and West's (1998b) proposal that normative responding on a variety of critical thinking tasks was linked, to explore the relationships between performance on the tasks and individual differences in cognitive ability and thinking dispositions, and to replicate performance patterns which have been argued in the literature to have been influenced by heuristic and analytic factors. The purpose of the study was also to develop the tasks and individual differences measures that will be used in the following training experiments.

There was some evidence to suggest that performance on the tasks was related. Individuals who utilised more analytic reasoning strategies on the experiment evaluation problems, that is individuals who identified the design flaws, were more able to reason analytically on the arbitrary selection tasks and achieve the correct response, achieved a higher number of correct responses on syllogistic reasoning problems involving conflict between logic and belief, and relied more on objectivity when evaluating arguments. Participants who relied more on objectivity also performed better on syllogisms involving conflict and were more able to solve deontic reasoning problems. Law of large numbers reasoning was associated with performance on conflict problems only. Further relationships between the tasks were expected however the absence of a correlation is not conclusive. Due to the low number of participants in this study, this is more likely to be a function of low statistical power. It must be noted that in Stanovich's work the sample sizes are invariably in the hundreds.

No associations between the biases elicited on the different tasks were found. In accord with Stanovich and West's proposals that departures from normative responses on the critical thinking tasks were due to systematic limitations in processing, it was predicted that biases elicited on the tasks would be associated with each other. However the prediction was not supported. In addition to low power there are a number of possible explanations for this. Under a dual process account of reasoning, biased responses are cued by System 1 processes. These processes are highly contextualised and domain-specific therefore it would not be expected that bias on different tasks was related. Each task would elicit bias from different domains.

Next we investigated the relationships between performance on the tasks and individual differences in cognitive ability and thinking dispositions. Stanovich proposed that cognitive ability and thinking dispositions predict decontextualised reasoning performance in critical thinking, whereas Klaczynski et al. argued that the two factors are independent. They reported that cognitive ability predicted sophistication of reasoning responses but decontextualised reasoning was predicted by the thinking styles measure of rationality. Two separate thinking styles measures were utilised: Stanovich and West's (1998b) Thinking Disposition Questionnaire and Epstein's (1994) Rational-Experiential Inventory both of which have been found to be reliable self-report measures here and in previous studies (Newstead et al., 2004; Handley et al., 2000).

The findings were by no means clear-cut and are difficult to interpret. Cognitive ability was associated with performance on the experiment evaluation problems and selection task performance. In other words, participants who identified more design flaws in the hypothetical experiments were higher in ability. The relationship between arbitrary task performance and ability was as predicted by Stanovich (1999). He proposed that it is the higher ability participants that are able to resist the heuristic (p/q) response on these tasks. The correlation with deontic tasks was almost of the same magnitude which could be due to the analytic responders on this task (Newstead et al., 2004). Both Newstead et al. (2004) and Klaczynski (2001) have found associations between ability and performance on deontic reasoning tasks.

Participants who reported that they were more rational as measured by the REI were better at performing on syllogistic reasoning tasks involving conflict between logic and belief. However a surprising result was the correlations between biased responding on the experiment evaluation problems and rationality (and Stanovich's TDC). Participants who enjoy engaging in such logical reasoning tasks and report that they are open-minded flexible thinkers were more biased in their responses and in their ratings of the experiment strength and validity. A key problem with a self-report measure is that people may often opt for the more socially desirable response. Therefore people may like to think that they are more able and enjoy engaging in logical tasks, especially when they are in an experimental setting and know they are about to perform on reasoning problems.

Stanovich and West's Thinking Disposition Composite score yielded several associations. Responses on the TDC were associated with performance on syllogistic problems involving conflict and no conflict between logic and belief. In other words participants who reported that they were more open-minded, flexible thinkers were more able to reason successfully on problems where logic and belief cue the same response and where they cued conflicting results. TDC responses were also associated with arbitrary selection task performance as would be predicted by Stanovich. These participants also relied more on objectivity when responding on Stanovich's AET. Responses on the TDC were also negatively related to bias on the rating scales for law of large numbers reasoning problems. More open-minded participants were less biased when evaluating the strength and the persuasiveness of the arguments, which is consistent with Klaczynski's (1997; Klaczynski, Gordon, and Fauth, 1997) findings.

One of the main problems with these measures is that they are self-report questionnaires: they require participants to reflect on their own behaviours and feelings. The question is whether or not this is a valid way of measuring information processing styles as participants are being asked to think about how they think which is a metacognitive awareness that individuals may not have. However, Klaczynski, Gordon, and Fauth and Klaczynski (1997) have found associations between information processing styles and biases in reasoning. Additionally, Pacini and Epstein (1994) have shown relationships between the latest version of the REI and aspects of personality. Therefore several studies have already provided evidence for the validity of such a measure.

One outstanding issue is that any associations that have been found with reasoning are on the Rational scale. There is no evidence yet to date of any significant relationships between the Experiential scale and any other measures of individual differences. It would also be expected that experientiality would be associated with heuristic responding on reasoning tasks. If it were a measure of the tendency to rely on System 1 processing then it would be expected that it would be associated with performance on deontic selection tasks or goal-consistent critical thinking tasks. The results here are supportive of Newstead et al.'s conjecture that the experiential scale is not a measure of System 1 thinking. It does appear that you cannot measure unconscious processes using a self-report questionnaire.

Klaczynski's statistical reasoning and experiment evaluation reasoning performance findings were partially replicated, but not very convincingly. Participants changed their strategies dependent on whether the conclusion presented was enhancing, threatening or neutral to an individual's occupational goal. More sophisticated reasoning strategies were utilised on problems involving goal-threatening conclusions than goal-enhancing or neutral conclusions but the effects were either very small or not significant. In Klaczynski's terms, individuals are cognitive misers who expend little effort processing non-threatening evidence and information. When the evidence is threatening to an individual's goal or belief, then dissonance is created and the individual has to expend greater processing effort to reduce it. Strategies such as the law of large numbers, which are usually difficult to access, are activated by the additional cognitive expenditure. When the evidence is not threatening, because it is consistent with the individual's goals or beliefs, it is easily assimilated to pre-existing belief systems and processed with relatively little effort. Therefore less sophisticated heuristic strategies are utilised (Klaczynski & Fauth, 1997).

However, we failed to convincingly replicate these findings. One explanation is that the occupational goal manipulation was less effective for our population than for the American population in Klaczynski's studies. Most of the participants in this study were first year psychology undergraduates. The occupations that they indicated as their goal may not have been an ambition they felt that passionate about or may have just been a momentary desire. In which case they would not have been so driven to defend their beliefs about the goal in question and also may not have had much of a belief system about that occupation to defend in the first place.

Cognitive ability was not found to be associated with law of large numbers reasoning performance; however it was associated with more sophisticated reasoning strategies on the experiment evaluation problems. In other words, individuals with higher ability are more able to detect the flaws in the hypothetical research. Klaczynski (1997) found that cognitive competence was present in his participants but use of it was selective. Only when challenged by the situation did they produce algorithmic solutions to problems i.e. theory-threatening arguments. According to Klaczynski, the biases elicited by the different problems are not due to ability but due to thinking styles. There is some support for that here as the more open-minded, flexible thinkers, as measured by the TDC are less biased when rating the strength and persuasiveness of the statistical reasoning arguments. However, participants who report to be more rational and open-minded are also more biased in their responding on the experiment evaluation problems.

Klaczynski has replicated his findings using different measures, music theories questionnaires (Klaczynski, 1997) and occupational goals (Klaczynski, Gordon, & Fauth, 1997; Klaczynski & Fauth, 1997) therefore it is possible that the measure does not elicit the same strength of belief here as it does in the United States. To obtain stronger effects it may be necessary to design a different set of materials based on the ones used in this study which will elicit more strongly held implicit beliefs such as stereotypes. Another reason for failing to find larger significant differences between reasoning strategies on these measures is that participants in this study were asked to complete a variety of tasks over the two sessions which may have resulted in fatigue. Therefore participants were not performing at their optimum level when requested to write explanations for the eighteen problems in the two everyday reasoning tasks.

This study provided support for the reliability and validity of Stanovich and West's (1997) Argument Evaluation Test. The task successfully identified two different groups of participants: one group who relied more on objective argument quality for argument evaluation decisions and the other who relied more on their prior beliefs. No associations were found between objective argument quality and ability as Stanovich and West reported. However objectivity was related to resolution of conflict problems in syllogistic reasoning, the identification of design flaws in hypothetical experiments, and performance on deontic selection task problems. These findings provide support for Stanovich and West's proposal that the AET is a measure of the ability to reason in situations in which prior beliefs may be interfering. Again a more powerful study might have resulted in stronger associations being found.

The typical effects from the syllogistic reasoning literature were found. More valid conclusions were endorsed than invalid and more believable conclusions were endorsed than unbelievable, irrespective of their logical validity (Evans, Barston, & Pollard, 1983; Newstead, Pollard, Evans, & Allen, 1992). No associations were found with cognitive ability and only a couple with thinking styles. More specifically, participants who report themselves to be open-minded rational thinkers appear to be more able to correctly solve problems where belief and validity are in conflict i.e. valid unbelievable and invalid believable items and where they are not in conflict i.e. valid believable and invalid unbelievable items. This is slightly in contrast with Newstead et al. (2004) who found ability was related to performance on conflict problems and thinking style was associated with performance on nonconflict problems. The relationship here with conflict problems was only small and as Newstead et al. proposed the rationality scale is really a measure of willingness to engage in thinking tasks. Therefore it is possible that the participants in this study were more motivated. However there was no relationship with cognitive ability to support Newstead et al.

The pattern of results on performance on the selection tasks was consistent with previous findings in the literature. Regardless of the theory being proposed, pragmatic reasoning schemas (Cheng & Holyoak, 1985), a Bayesian model of optimal data selection (Oaksford & Chater, 1994) or social contracts (Cosmides, 1989) performance on the selection tasks was facilitated by deontic content compared to performance on the abstract versions. Consistent with Stanovich and West's (1998a; 1998b) findings, there was a small association between cognitive ability and correct responding on the abstract versions.

The tasks used were the same as those used by Newstead et al. who found only moderate test-retest reliability correlations on accuracy for the abstract and deontic tasks (.62 and .34 respectively, overall .38) suggesting random variation in performance on the selection tasks. For the next study involving selection tasks additional problems of each type will be used in order to obtain more reliable data.

3.5 Conclusions

The first aim of this study was to study the relationships between performance on all the reasoning and decision making tasks. Stanovich (1998b) claimed that normative responding on a range of critical thinking tasks was related. The pattern of correlations illustrated in this study provides some evidence to support this even though the sample used was small for an individual differences study.

A second purpose of this study was to examine the relationships between performance on the tasks and measures of individual differences in thinking styles and cognitive ability. Stanovich proposed that thinking dispositions and ability predict performance on critical reasoning tasks. In contrast, Klaczynski argued that ability predicts sophistication of reasoning responses and bias is predicted by thinking style. The correlations reported here are not strong enough to support either argument. There were very few associations with ability and the thinking styles measures yielded a few relationships with both logical and biased performance patterns on the tasks.

One of the main aims of this study was to develop the materials for use in the experiments reported in the following chapters. Most of the robust effects reported in the literature have been found, even with a relatively small number of participants. Klaczynski's findings in relation to belief-based responding on everyday reasoning and experiment evaluation problems were only partially replicated. These findings were not as strong as expected therefore it is necessary to attempt to replicate them using a different set of materials. The aim of the next series of experiments reported in this thesis is to investigate the effects of training and instruction on the different types of responses elicited by the belief-motivated reasoning tasks. A clearer distinction may be found between heuristic and analytic responding on these tasks if problems involving more strongly held beliefs such as social stereotypes are utilised. Therefore new materials based on the ones used in this study will be designed.

Selection task performance was as expected. Problems consisting of arbitrary content yielded a much lower solution rate than problems consisting of deontic content. The relationships with ability were small however by employing a higher number of selection tasks of each type for the training experiments will provide a more reliable measure of pragmatic and analytic selection task responding. In line with the training study on belief-motivated reasoning, in Chapter 5 the aim is to investigate the impact of training on the different types of responding in selection task reasoning. It will only be possible to interpret findings under a dual process account if there are clear distinctions between responses under the two systems.

In the next chapter a series of studies are reported which explore the effects of training in the concept of the law of large numbers on a variety of everyday reasoning tasks.

CHAPTER 4

Effects of Training in Statistical Principles on Reasoning and Bias

4.1 Introduction to Experiment 2

The purpose of the experiments reported in this chapter is to examine the effects of training in statistical principles on law of large numbers reasoning and bias. As discussed in Chapter 1, there has been a great deal of research conducted into the pragmatic and analytic influences on reasoning performance. It is becoming more widely recognised that theoretical accounts of reasoning and decision making must explain the influences of both analytic and heuristic processes on people's responses. There have also been a variety of training studies, as discussed in Chapter 2, designed to investigate transfer of knowledge and skills deemed necessary for optimum performance on reasoning and decision making tasks. Apart from the belief bias in syllogistic reasoning studies, most of these studies have concentrated on the improvement of analytic processes only and failed to observe the effects these training manipulations may have on belief-based processes. In the two experiments reported here, the aim is to investigate the extent to which training in an inferential rule system, the law of large numbers' impacts on reasoning performance. Further, to examine the extent to which it impacts on the belief-based and analytic processes which influence everyday reasoning competence.

This research brings together two bodies of research: the statistical training studies conducted by Fong, Krantz, and Nisbett (1986; Fong & Nisbett, 1991) and more recently Klaczynski's work exploring individual differences in statistical reasoning and bias across the lifespan (Klaczynski, 1997; Klaczynski & Fauth, 1997; Klaczynski & Gordon, 1996; Klaczynski, Gordon, & Fauth, 1997; Klaczynski & Robinson, 2000). The experiments have been designed to examine the effects of explicit, rule-based training and instruction on belief-based and analytic reasoning processes utilised on everyday inferential problems.

A key ability that a person requires for effective critical thinking is the ability to evaluate evidence independently from their goals and beliefs (Klaczynski, Gordon, & Fauth, 1997). Klaczynski and colleagues have conducted a number of studies investigating the relationships between critical reasoning, intellectual ability and information-processing styles (Klaczynski & Robinson, 2000; Klaczynski, 1997; Klaczynski & Fauth, 1997; Klaczynski, Gordon, & Fauth, 1997; Klaczynski & Gordon, 1986). Klaczynski, Gordon, and Fauth identified three critical thinking competencies that people may utilise. These were law of large numbers reasoning, 'intuitive analyses of covariance' (the ability to cognitively control for third variable effects that may underlie observed relationships) and the ability to dissect hypothetical experiments for flaws. Klaczynski et al. found that participants utilised more sophisticated reasoning strategies in all three types of problems when information presented to them was contrary to their beliefs. When the information was consistent with their beliefs it was processed in a relatively cursory way.

It has been shown repeatedly that when presented with evidence that is contrary to a person's beliefs, more sophisticated reasoning strategies will be utilised to discredit the evidence (Lord, Ross, & Lepper, 1979). Klaczynski and Fauth (1997) gave participants nine everyday reasoning problems designed to elicit law of large numbers reasoning and nine experiment evaluation problems. Three involved evidence or arguments that were enhancing to the participant's occupational goal, three were threatening and three neutral. In support of previous research, problems that involved evidence which was threatening to their goals elicited more sophisticated reasoning strategies. In the case of LLN problems people utilised LLN reasoning and for the experiment evaluation problems people searched for the confounds which would make the experiment invalid. Klaczynski et al. concluded that strategies are changed to suit the goal of the individual at an intrinsic level and the biases participants are displaying are self-serving.

Why do problems involving information or conclusions that are incongruent with one's beliefs elicit more sophisticated reasoning strategies than problems involving congruent or neutral information? There are several explanations that have been proposed to account for such findings over the years. Tversky and Kahneman (1973) introduced the 'availability heuristic' to account for biased responding on judgement and probability tasks. It was argued that the ease with which examples could be 'brought to mind' influenced responding. Some things are brought to mind easier than others resulting in reasoning bias. Tversky and Kahneman also presented data that illustrated that bias arises from selective encoding and retrieval of evidence which supports a prior belief, therefore leading to maintenance of false theories or self-serving reasoning biases (Klaczynski & Fauth, 1997).

Related to this is the notion of 'confirmation bias', which is the tendency for people to seek information that is consistent with their beliefs or theories and to avoid the collection of potentially falsifying evidence. Evans (1989) viewed confirmation bias not as a motivational bias that maintained belief structures but as a result of cognitive failures and referred to it as a 'positivity bias'. Positivity bias was seen to reflect the operation of pre-conscious processes which direct attention to positive rather than negative information. According to Evans (1989) the major cause of bias in human reasoning and judgement are factors which induce people to process the problem information in a selective manner. If logically relevant information is excluded during problem representation or logically irrelevant information is included then bias will result.

Support for this has also been found in the belief bias in syllogistic reasoning literature. The Selective Scrutiny account of belief bias claims that people focus on the conclusion and only engage in logical processing if this is unbelievable. Representational heuristics precede logical analysis and cause the analysis to be applied to certain aspects of the problem content, therefore resulting in a lack of logical analysis on problems involving believable conclusions (Newstead, Pollard, Evans, & Allen, 1992). Oakhill, Johnson-Laird, and Garnham (1989) propose a mental model theory of belief bias and claim that when the conclusion conflicts with belief, participants are motivated to search for counter-examples. This still doesn't answer why people use such strategies. Nisbett and Ross (1980) suggest that people tend to persevere in their beliefs well beyond the point at which logical and evidential considerations can sustain them. They suggest that this belief perseverance seems to occur sometimes because people have an emotional commitment to the belief, a proposal supported by Klaczynski (1997) who reported on the self-serving nature of reasoning biases in adolescents.

In terms of describing the belief and analytic influences on these tasks, Klaczynski, Gordon, and Fauth (1997) propose a depth of processing explanation for the moment-to-moment shifts in reasoning behaviour. Cognitive-Experiential Self Theory is a dual-process theory which claims that reasoning involves using two parallel, independent systems; the rational and the experiential systems. It is suggested that reasoning is an interaction between the two and in the case of being presented with evidence that is contrary to belief, an individual's analytic system (rational system) is triggered which results in information being processed at a deeper level. This involves more cognitive expenditure but results in the activation of more sophisticated reasoning strategies such as the law of large numbers, therefore allowing the individual to refute the evidence that is inconsistent with their beliefs. When the evidence is consistent with the beliefs of the individual, then the information is processed at a shallow level by the experiential system. The evidence is assimilated to the already pre-existing beliefs and the conclusion accepted with little or no cognitive expenditure. In their view, sophistication of reasoning responses is related to ability. However, degree of bias is related to thinking style. The aim of these experiments is to investigate whether individuals can be taught or instructed to utilise the same strategies on all the problems, regardless of the direction of belief-laden content.

As we have seen, Fong, Krantz, and Nisbett (1986) conducted a series of experiments to investigate law of large numbers reasoning. In a series of experiments, participants were either trained in the LLN rule system which consisted of a description of the concept of sampling and the law of large numbers; given examples training which consisted of three problems in a given domain i.e. probabilistic, objective or subjective, followed by an explanation of how to solve the problem in LLN terms; or Full training which consisted of Rules training followed by Examples training. Participants were then given everyday reasoning problems in the probabilistic, objective and subjective domains. Previously it had been found that problems involving probabilistic content evoked LLN reasoning the most (Krantz & Nisbett, 1983).

Fong et al. found that both the rule training and examples training improved statistical reasoning and enhanced the quality of the reasoning for problems across all three domains. The rule training plus examples was found to have an additional effect. The studies did indeed provide evidence for the domain-independence of training. Participants who were given examples only training using problems from the objective domain were able to utilise the statistical principles across all three problem domains on testing. Further evidence was provided in their following experiment (Exp. 2) when they trained their participants in one of the three domains and found that the training significantly increased use of statistical reasoning regardless of the domain of training. From the results Fong et al. concluded that participants were able to map the LLN rules they had learnt onto a pre-existing set of abstract intuitive rules that they then used on problems in different domains to the one that they had been taught.

It could be argued that the domains used in the original study were too broad and there were no domain-independence effects of training at all. Fong and Nisbett (1991) used the more tightly defined domains of sports and ability testing and found use of statistical principles was still improved after a two-week delay, however domain specificity of training was observed. There was a loss of training effects over the delay in the untrained domain (either sports or ability testing) but these participants still applied the LLN heuristic more than participants from the untrained control group.

One of the key questions that the present research would like to address is would training in an inferential rule system transfer to problems involving belief-laden content? Is there a way of getting people to utilise their analytic systems on the belief-consistent as well as belief-inconsistent evidence? According to dual process theories, belief-based influences on reasoning are not the result of conscious thinking but System 1 processes. These pre-conscious processes cue belief-based responses which impact on reasoning and decision making behaviours automatically. However it is System 2, the analytic system that is sensitive to instruction and explicit training. Instructing people on the rule may give them the analytic strategies they need to override the 'fundamental computational bias' (Stanovich, 1999) of System 1. According to the research performed on belief bias in syllogistic reasoning reviewed in Chapter 2, belief bias may be reduced but not eliminated by the manipulation of instructional procedures (Newstead et al., 1992; Evans, Newstead, Allen, & Pollard, 1994). This research did however illustrate that people do have some degree of conscious control over their logical reasoning processes.

4.1.1 Rationale for Experiment 2

The main aim of Experiment 2 was to observe the effects of explicit training in statistical principles on reasoning and bias. This is the first time that this training has been tested on belief-laden materials. Belief bias was reduced by instructional manipulation in the deductive reasoning literature (Evans, Newstead, Allen, & Pollard, 1994; Newstead, Pollard, Evans, & Allen, 1992). Evans et al. found that the bias was reduced by emphasising the structure of the syllogism, explaining the logical meaning of 'SOME' and also emphasising the need to base responses only on the information given. In particular, they reported a reduction in the acceptance of invalid-believable arguments.

Belief bias effects are noted by the much higher acceptance rates of believable rather than unbelievable conclusions. The belief bias effect is more marked on invalid problems. That is people will readily endorse it as valid due to its believability. This is in line with participants' performance on the everyday reasoning problems. When they are given a problem that is consistent with their beliefs they process the information at a very cursory level and accept the conclusion without searching for the flaws in the argument.

The findings in relation to performance after the instructional manipulation have been interpreted under a dual process account. This account attributes System 2 and 1 processes to the logical and belief-based processes respectively that are influencing the task (Evans, 2003). Instruction was found to reduce bias on these tasks and with the assumption that System 2 processing inhibited the automatic System 1 processes.

By bringing together the two major bodies of research, Fong et al.'s training studies and Klaczynski's individual differences research, it will be possible to investigate whether the same pattern of findings may be obtained on a different type of reasoning task. Participants will be presented with everyday reasoning problems as used by Fong et al. (1986), plus experiment evaluation and law of large numbers reasoning problems involving belief-laden content such as those used by Klaczynski, Gordon, and Fauth (1997). Participants in the training condition will be given training on the concept of the law of large numbers.

It is predicted that in the absence of training, more law of large numbers reasoning will be utilised on problems consisting of probabilistic content, followed by those involving objective content. Subjective problems will elicit the least LLN reasoning. On problems involving belief-laden content, participants who receive no training will utilise more sophisticated reasoning strategies on problems consisting of evidence or conclusions which are inconsistent with a person's beliefs. Law of large numbers reasoning will be utilised more on statistical reasoning problems and more validity confounds will be identified in the experiment evaluation problems.

The predictions in relation to the problems involving belief manipulations after training are not so clear. Taking the findings in the belief bias literature, bias may be reduced after training. Training will impact on System 2 and a function of this will be to override or inhibit the belief-based responses cued by System 1. However, according to Klaczynski, the effects of belief and the level of LLN reasoning are independent and associated with different systems. Training (if it has any effect at all using these materials) will impact on the amount of statistical reasoning utilised but it will have no impact on the bias. Hence System 2 instruction may impact only on the level of LLN reasoning, but not belief.

Prior to Experiment 2, a small-scale pilot study was conducted to ascertain typical, untypical and neutral character traits or behaviours for different professions and occupations. In their studies, Klaczynski, Gordon, and Fauth ascertained participants' occupational goals and utilised them in the everyday reasoning problems to attain goal-biased reasoning. This resulted in a very lengthy and complicated process of finding out each participant's goal in one experimental session and then rewording each problem to suit each individual for the next. In Experiment 1 of this thesis, Klaczynski et al.'s methodology was repeated; however the manipulation failed to find strong effects on the goal-threatening materials.

For the following statistical training studies a new set of materials are required which strongly engage both belief-based and analytic reasoning strategies. In dual process terms, belief-consistent arguments are processed at a cursory level by System 1 and belief-inconsistent arguments are processed using more sophisticated strategies by System 2. Using such materials would enable us to investigate the extent to which the training impacts on both types of reasoning processes. Also, new materials would possibly make the experimental procedure easier. In addition if Klaczynski et al.'s results were replicated using novel problems, it would demonstrate the generalisability of their findings.

4.2 PILOT STUDY

4.2.1 Method

Design

Participants were administered a questionnaire involving thirty-four different professions and occupations. The aim was to determine typical, untypical and neutral character traits for the different occupations that could be utilised in the everyday reasoning and experiment evaluation problems in Experiment 2.

Participants

18 participants from the University of Plymouth, undergraduates and postgraduates, took part in the pilot study.

Materials

The inventory consisted of 34 familiar occupations and professions. Following each occupation was a list of six traits or behaviours that may or may not be typical of people in that profession (see Example 4.1 below for an example of an item). Participants were required to indicate on a scale of 1 to 5 (1 being very untypical, 5 being very typical) how typical they rated each trait/behaviour for each particular occupation (See Appendix A2.1 for full inventory).

Occupation	Traits	Untypical				Typical
Nurses are	caring	1	2	3	4	5
	aggressive	1	2	3	4	5
	thoughtful	1	2	3	4	5
	intelligent	1	2	3	4	5
	lazy	1	2	3	4	5
	healthy	1	2	3	4	5

Example 4.1 Example of an item taken from the occupation/traits inventory

Procedure

Participants completed the inventory within their own time limits.

4.2.2 Results

Descriptive statistics were obtained for the scores on each trait/behaviour related to each occupation. The mean scores were then sorted with the highest mean score first and the lowest mean score for a trait last (see Appendix A2.2). The six highest and six lowest character traits (most typical and least typical related to professions) could then be identified. Six neutral traits (neither typical nor untypical) were also identified. A repeated measures ANOVA (typical x untypical x neutral) was performed which illustrated that the mean ratings for each set of professions/traits were significantly different from each other (at $F(2, 34) = 327.526$, $MSE = 42.5067$, $p<.001$; all p less than .001). These items were then used to design the Law of Large Numbers problems and the Experiment Evaluation problems in Experiment 2. See Appendix A2.3 for the descriptive statistics of the items identified.

4.3 Method - Experiment 2

Design

The experiment was designed to partly replicate Fong, Krantz and Nisbett (1986, 1991) and extend on the work of Klaczynski (Klaczynski & Robinson, 2000; Klaczynski, 1997; Klaczynski & Fauth, 1997; Klaczynski, Gordon & Fauth, 1997). A between subjects design was used involving two conditions, Control and Training. The Control group received no training whereas the participants under the Training condition received Fong et al.'s full training. That is training on the concept of the law of large numbers followed by examples of everyday reasoning problems. Each example was followed by a full explanation of the answer in terms of LLN reasoning.

Participants

The participants were 60 undergraduates from the University of Plymouth, 51 female and 9 male (mean age 22.1, stdev. 5.84) who were taking part for course accreditation. Groups of participants were randomly allocated to each condition, resulting in 30 taking part under each one (mean age Control group = 24.63, stdev. 7.36; Training group = 19.56, stdev. 1.4).

Materials

Instructions:

The instructions for both conditions were taken from Fong et al. (1986). The instructions for the Control group read:-

We are interested in studying how people go about explaining and predicting events under conditions of very limited information about the events. It seems to us to be important to study how people explain and predict under these conditions because they occur very frequently in the real world. Indeed, we often have to make important decisions based on such explanations and predictions, either because there is too little time to get additional information or because it is simply unavailable.

On the pages that follow, there are a number of problems that we would like you to consider. As you will see, they represent a wide range of real-life situations. We would like you to think carefully about each problem, and then write down answers that are sensible to you.

These were presented prior to the Test Materials booklet. Participants in the Training condition were presented with the first paragraph of the above prior to the training itself. The second paragraph was presented prior to the test materials, and ended in the sentence, "In many of the problems, you may find that the Law of Large Numbers is helpful."

Full Training

The training started with the first paragraph of the instructions described above, followed by a paragraph introducing the law of large numbers:

Experts who study human inference have found that principles of probability are helpful in explaining and predicting a great many events, especially under conditions of limited information. One such principle of probability that is particularly helpful is called the Law of Large Numbers.

Following this participants read a two-page description of the concept of sampling and the law of large numbers using examples of red and white beads in a jar, 30% red and 70% white. The beads in the jar represented the population, the proportion of red and white beads the population distribution and a selection of beads from the jar a sample. After the concept of sampling was explained (see Appendix A2.4 for full text plus demonstration script and training examples), the law of large numbers was presented:

As the size of a random sample increases, the sample distribution is more likely to get closer and closer to the population distribution. In other words, the larger the sample, the better it is as an estimate of the population.

Participants were given time to finish reading the description and then the experimenter gave a demonstration of the law of large numbers, using a jar containing red and white beads with the same population distribution as that of the written description - 70% white, 30% red. The experimenter stated the main concepts again (see Appendix A2.4 for full spoken script) and then proceeded to draw samples from the jar, four of size 1, four of size 4 and four of size 25, to demonstrate that the average deviation of a sample from the population would decrease as the sample size increases as the law of large numbers predicts. The experimenter and the participants summarised each sample on a table, keeping track of the deviation between each sample and the population.

Following the demonstration the participants were given a set of three example problems with an answer following each one that provided an analysis of it in terms of the law of large numbers. The examples were introduced by the following paragraph:

One reason that the law of large numbers is important to learn is that it applies not only to jars and beads. The basic principles involved in the law of large numbers apply whenever you make a generalisation or an inference from observing a sample of objects, actions or behaviours. To give you an idea of how broad the law of large numbers is, we have, in this booklet, presented three situations in which the law of large numbers applies. Each situation is analysed in terms of the law of large numbers.

The three example problems were taken from Fong et al. and consisted of a Structure 1 problem (generalising from a small sample), Structure 2 (regression - Structure 3 in Fong et al.) and Structure 3 (large sample vs. theory without supporting data - Structure 5 in Fong et al.), and were presented in that order. All three problems were objective in type. Fong et al. found that training on objective example problems improved performance on both probabilistic and subjective problems as much as it improved performance on the objective problems.

Participants were asked to read each one and then consider it for a few minutes before turning the page to read the law of large numbers answer (See Appendix A2.4 for the three example problems and their law of large numbers answers). The format of the answers was constant across each structure and included the following characteristics (Fong et al., 1986):

1. A statement about the goal of the problem.
2. Identification of the sample or samples and their distributions in the problem.
3. Explanation of how the law of large numbers could be applied to the problem. This identified the population distribution(s) and explained the relationship between the sample(s) and the population(s).
4. The conclusion that could be drawn from the application of the law of large numbers.

Test Materials:

Twenty seven problems were used in total. nine Law of Large Numbers and nine Experiment Evaluation problems adapted from Klaczynski, Gordon. and Fauth (1997). and nine problems taken from Fong et al. (1986).

Law of Large Numbers Problems:

As in Experiment 1, hypothetical individuals presented arguments and evidence that were either consistent or inconsistent with participants' beliefs, but instead of personal occupational goals, occupations and typical or untypical personality traits were used (see Pilot Study). Belief-consistent problems involved arguments for a positive correlation between an occupation and a typical personality trait (e.g. firemen are brave), whereas belief-inconsistent problems involved arguments for a correlation between an occupation and an untypical personality trait (e.g. firemen are cowards). Within each problem structure, there were three problem types, consistent (typical), inconsistent (untypical) and neutral (neither typical nor untypical). See Example 4.2 for an example of an argument resulting in a conclusion which involves a correlation between an occupation and an untypical personality trait, a belief-inconsistent problem (all problems are presented in Appendix A2.5).

An editorial in a local newspaper recently criticised the occupation of being an Aerobics instructor. The journalist's argument was:

I've got a friend who's an aerobic instructor and I wouldn't want anyone I know to copy her lifestyle. She is so unhealthy. She drinks and smokes and actually never takes any real exercise herself, she just tells others how to do so! I know her flatmates and they say she never stops eating as well, not healthy food either, fry ups and chocolate are normal. My conclusion? I don't think there's a more unhealthy group of people than Aerobics instructors!

Example 4.2 An example of a law of large numbers reasoning problem involving a belief-inconsistent argument

The nine problems consisted of 3 structures as in Experiment 1. Structure 1 problems involve hypothetical arguments in which an individual draws a conclusion from a single instance or from a very small sample of observations. Structure 2 problems involve presenting two hypothetical arguments. One individual's conclusion is based on a small sample and personal experience; a second individual draws a contradictory conclusion from a larger database. Structure 3 problems involve a hypothetical individual drawing a hasty conclusion from a small sample and from an outcome that was clearly deviant from what should have been expected.

'Persuasiveness' and 'Evidence Evaluation' Scales:

Following each problem, participants indicated on two 9-point scales how convinced they were by the argument (1=not at all convinced; 9=very convinced) and how strong they thought the conclusion was based on the evidence used (1=very weak; 9=very strong). These were then referred to as the 'persuasiveness' and 'evidence evaluation' ratings respectively. Total scores on each rating scale were calculated separately for belief-consistent, belief-inconsistent and belief-neutral problems. Scores could then range from 3 to 27 for each problem type. Participants were then required to write explanations of why the conclusions were convincing/not convincing and strong/not strong.

Coding of explanations:

These explanations were coded using the 3-point system used in Experiment 1, developed by Fong et al. A score of '0' was given if the response contained no indication of statistical reasoning. There was no mention of sample size, random variability, or the role of probability. For example, 'My friend is an aerobics instructor and she's healthy'. A score of '1' indicated that the participant referred to the law of large numbers, but vaguely. For instance, 'but that's just one aerobic instructor'. The participant implied that he or she was using statistical reasoning, but was not explicit about the statistical basis of his or her reasoning. If a participant scored '2' for a response, it meant that the LLN principle was clearly applied in the explanation, for example, 'the conclusion is a bad one as not all aerobics instructors are unhealthy, the journalist is only talking about one instructor. If he was to look at a larger sample of aerobics instructors, he may be able to report a more convincing argument'. For each problem type, scores were collapsed and added together to produce three total reasoning scores that range from 0-6.

The principal researcher and a second coder who was blind to condition and problem types independently coded all items for the 60 participants. Total agreement was achieved on 98% of the items. Agreement on the remaining 2% was achieved after discussion.

Experiment Evaluation Problems:

Participants were informed prior to the presentation of the experiment evaluation problems that they would be presented with several summaries of actual psychological and sociological research. Nine scenarios were developed, each containing a brief description of the hypothetical individuals who participated, their occupations, the methods used to conduct the research, and the findings and conclusions drawn from the research. The conclusions were belief-consistent, belief-inconsistent or belief-neutral as in Experiment 1 but involving occupations and typical or untypical personality traits. See Example 4.3 for an example of an experiment evaluation problem consisting of a belief-inconsistent conclusion (see Appendix A2.6 for all the problems within each problem type).

Dr. H. is a researcher who is interested in finding out how capable people in different occupations are of holding pleasant conversations with strangers. Dr. H. is especially interested in finding out how builders, teachers, solicitors and shop assistants are in dealing with strangers because people in these professions must deal with strangers on a daily basis. To observe their conversations, Dr. H. watched a large number of builders having a conversation with Jim, one of Dr. H's laboratory assistants. Next he watched the teachers, solicitors and shop assistants having a conversation with Amy, another one of his research assistants. Dr. H. noted that teachers, solicitors and shop assistants overall made more inappropriate hand gestures, more body contact with the other person and more wandering eye movements than the builders. He concluded as part of his journal article that builders were more politically correct than any of the other participants from other occupations.

Example 4.3 An example of an experiment evaluation reasoning problem involving a belief-inconsistent argument

The three problem structures described in Experiment 1 (Klaczynski et al., 1997) were utilised again. Structure 1 problems involved descriptions of pseudo-experimental research in which the primary independent variable was confounded with another variable. Structure 2 problems described research comparing the occupations, however, a selection confound was built into the research such that the occupational groups were selected from very different populations. In Structure 3 problems, the construct validity of the dependent variable was suspect. Within each problem structure there were three problems, one belief-consistent (typical personality traits correlated with occupation), one belief-inconsistent (untypical personality traits correlated with occupation), and one belief-neutral problem (neither typical nor untypical personality traits correlated with occupation).

'Conclusion strength' and 'Experiment validity' Scales:

Following each problem, participants rated the strength of the researcher's conclusion on a 9-point scale (1=extremely weak, 9=extremely strong), and the validity of the experiment (1=extremely invalid, 9=extremely valid). These were then known as the 'Conclusion strength' and 'Experiment validity' scales respectively. Then participants wrote explanations of why they believed the experiment to be weak/strong and invalid/valid. Scores on the two scales were collapsed across problem type and ranged between 3 and 27 for each.

Coding of explanations:

The coding system for the experiment evaluation problems was used in Experiment 1 and was based on a 3 point scale as used by Klaczynski, Gordon, and Fauth (1997). A score of '0' was given if the participant gave no indication that any threat to the validity of the experiment existed. For example, 'A lot of builders are more politically correct these days. There are rules now and they don't want to lose their jobs'. In other words the participant believed that the experiment was validly conducted or rejects the experimental evidence by simply asserting that it is true or false. A participant scored '1' if he or she indicated their awareness of the confound built into an experiment but did not indicate that the existence of this confound made it impossible to make a straightforward interpretation of the findings. For instance, 'the builders spoke to a different person which may have influenced the results a little'. Finally a score of '2' was awarded when a participant indicated that the experimental confound made it impossible to interpret the findings of and draw conclusions from the research, e.g. 'The builders spoke to a male whereas the other professions spoke to a female so the builders were bound to appear more politically correct. This does not mean they are. This experiment is invalid. Everyone needs to speak to the same person for the findings to be valid'.

All items for the 60 participants were coded by the principal researcher and a second coder who was blind to condition and problem types independently. Total agreement was achieved on 95% of the items. Agreement on the remaining 5% was obtained after discussion.

Fong et al.'s Everyday Reasoning Problems:

The nine problems were taken directly from Fong et al. (1986, Experiment 1) and consisted of three types as discussed in the introduction, probabilistic, objective and subjective. For the probabilistic problems, participants were asked to draw conclusions about the characteristics of a population from sample data generated in a way that clearly incorporated random variation. Randomness was made explicit in some way. In objective problems, participants had to draw conclusions about characteristics of a population on the basis of 'objective' sample data but with no explicit cue about randomness of the data. For subjective problems participants drew conclusions from subjective characteristics of a population from 'subjective' sample data.

Within each problem type, three problem structures were used (Fong et al. used six). Structure One problems required participants to draw conclusions from a single small sample. Structure Two (Fong's structure three) problems required participants to explain why an outcome selected because of its extreme deviation was not maintained in a subsequent sample. Structure Three (Fong's structure five) pitted a large sample against a plausible theory that was not founded on data. Therefore, resulting in 9 problems with problem type crossed with problem structure (see Appendix A2.7 for all problems).

All 27 test problems were presented in random order for each participant, with the constraint that no two problems with the same structure appeared successively.

Coding of explanations:

The 3-point coding system developed by Fong et al. was used (see LLN coding). The principal researcher and another coder who was blind to condition and problem types independently coded all items for the 60 participants. Total agreement was achieved on 96% of the items. Agreement on the remaining 4% was obtained after discussion.

Procedure

Participants took part in the experiment in groups of 2 to 6. The Control group were presented with the instructions as above and then told to proceed through the test material booklet.

Participants in the Training group were required to attend two sessions. In session one they received full training (as described above) and in session two, approximately one week later, they completed the test materials booklet. The training session took 40 minutes and the participants were allowed up to 1 hour 30 minutes to complete the test materials booklet.

4.4 Results

4.4.1 Law of Large Numbers reasoning

The first analysis to be presented will focus on the belief-laden law of large numbers reasoning problems. As described in the method section, subjects' responses on the nine LLN problems were coded using the 3-point system devised originally by Fong et al. (1986). A code of '0' was given for responses that contained no mention of statistical concepts, whereas a '1' or '2' was given for responses that involved statistical notions, '1' for a poor statistical response, '2' for a good statistical response. Scores were then collapsed across the three problem types, Neutral (problems with conclusions that were neither typical nor untypical of a person's prior beliefs), Consistent (problems with conclusions that were typical of a person's prior beliefs) and Inconsistent (problems with conclusions that were untypical of a person's prior beliefs).

As may be seen in Table 4.1 the level of statistical responding was higher for the participants that received training on the LLN principle on all three types of problems. It is also evident that law of large numbers reasoning is higher on responses involving belief-inconsistent conclusions than on belief-neutral or belief-consistent problems.

	LLN Neutral	St.dev.	LLN Consistent	St.dev.	LLN Inconsistent	St.dev
control	1.27	0.91	1.53	1.07	2.33	1.15
training	3.30	1.39	3.00	1.39	3.37	1.35

Table 4.1 Mean scores for Law of Large Numbers Reasoning on each type of problem under each condition (N =60: max. score = 6)

A 2 (Condition) x 3 (Problem Type) mixed ANOVA was performed with training as a between-subjects factor and problem type as a within-subject factor. Main effects of Condition ($F(1, 58) = 39.54$, $MSE = 102.76$, $p<.001$) and Problem Type ($F(2, 116) = 6.94$, MSE = 6.62, p<.01) illustrate that participants utilise more sophisticated reasoning techniques on all problems after training and LLN is utilised on problems that involve conclusions which are inconsistent with prior belief. An LSD follow-up test revealed reasoning scores on problems involving inconsistent information to be higher than scores on consistent or neutral problems (both $p < .01$). The ANOVA also yielded a significant interaction between Condition and Problem Type ($F(2,116) = 3.96$, $MSE = 3.77$, $p<.05$. see Figure 4.1 for graph of the interaction).

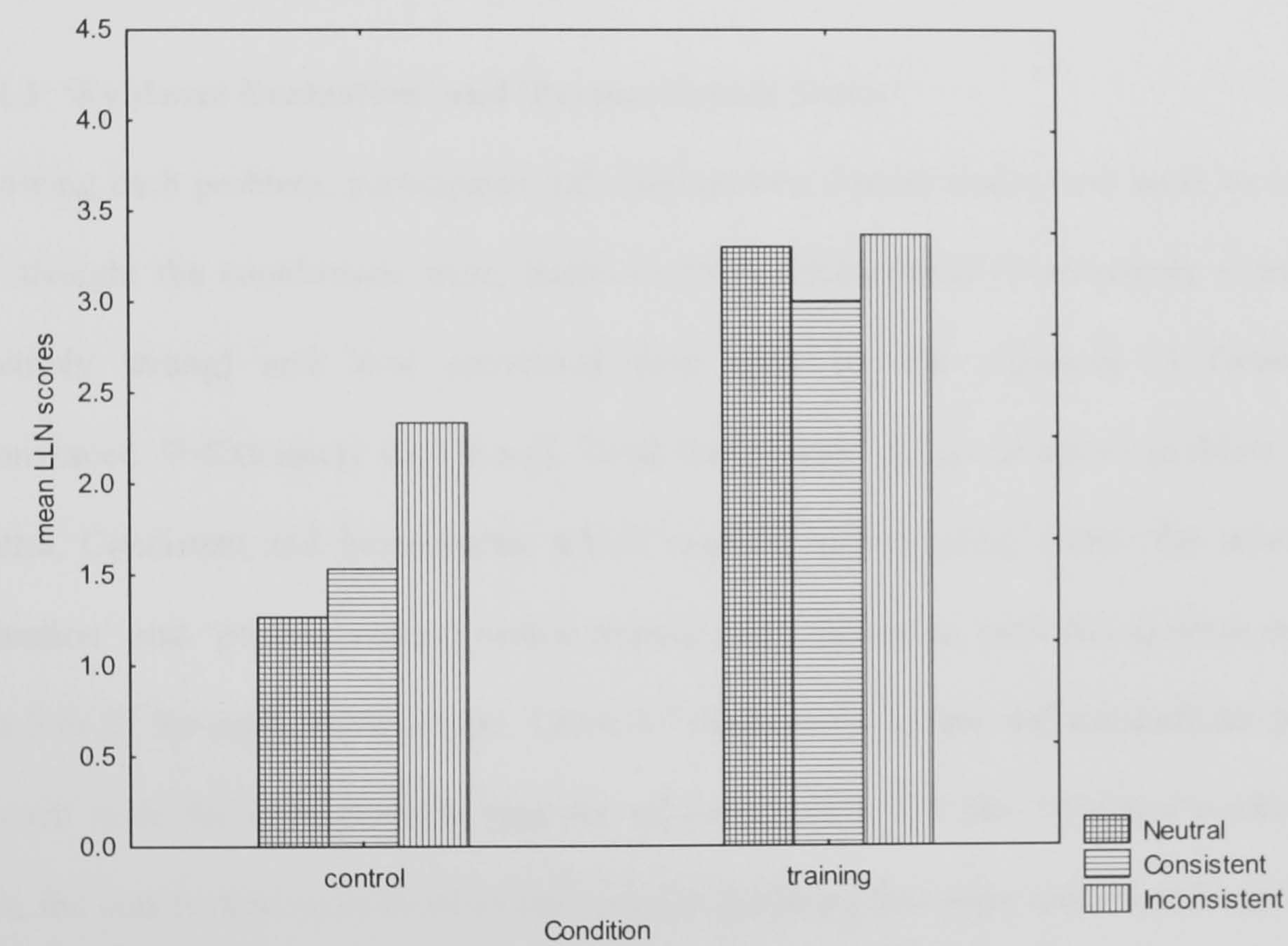

Figure 4.1 Interaction between Condition and Problem Type

As may be seen in Figure 4.1, LLN reasoning is utilised more for inconsistent problems than either consistent or neutral problems in the Control condition ($F(1, 58) = 8.9$, $MSE = 9.6$, $p<.01$ and $F(1,58) = 17.72$, $MSE = 17.07$, $p<.0001$ respectively), whereas with training use of LLN reasoning is improved greatly on all three problem types. There are no differences in sophistication of responses between any of the problem types ($F(1, 58) = 1.63$, $MSE = 1.35$, $p>.1$ between neutral and consistent; $F(1, 58) = 1.89$, $MSE = 2.02$, $p>.1$ between consistent and inconsistent; $F(1, 58) = 0.07$, $MSE = 0.07$, $p>.1$ In other words, when presented with problems or arguments that are inconsistent with one's prior beliefs, one will utilise a more sophisticated reasoning style to argue with, whereas if the information is consistent or neutral to one's beliefs then it is much less likely to be evoked. However, with training in statistical principles, statistical reasoning is more likely to be used whatever the problem type.

4.4.1.1 'Evidence Evaluation' and 'Persuasiveness Scales'

Following each problem, participants indicated on two 9-point scales how weak or strong they thought the conclusions were, based on the evidence used (1=extremely weak, 9= extremely strong) and how convinced they were by the argument (1=Extremely unconvinced, 9=Extremely convinced). Total scores were collapsed across problem type, Neutral, Consistent and Inconsistent, which resulted in two rating scales, the 'evidence evaluation' and 'persuasiveness' scales respectively. Scores on each rating scale ranged from 3 to 27 for each problem type. Table 4.2 displays the means and standard deviations for each scale for each problem type for all participants. For the 'evidence evaluation' scale, the conclusions were rated as stronger for problems that were neutral and consistent with a person's beliefs.

Scale	Problem Type	Control		Training	
		Mean	St.dev.	Mean	St.dev.
Evidence Evaluation	Neutral	12.53	2.73	14.53	3.14
	Consistent	14.10	3.90	13.33	3.92
	Inconsistent	8.77	3.77	9.40	3.09
Persuasiveness	Neutral	11.97	2.94	13.37	2.52
	Consistent	13.57	3.87	12.80	3.28
	Inconsistent	8.07	3.43	8.53	2.55

Table 4.2 Means and Standard Deviations on the 'Evidence Evaluation' and 'Persuasiveness' scales under both conditions (N=30 per condition).

A 2 (Condition) x 3 (Problem Type) mixed ANOVA with Training as a between subjects variable and Problem Type as within subjects found no effect of condition ($F(1, 58) = 1.02$, $MSE = 20.67$, $p>.05$) on the evidence evaluation scale. A main effect of Problem Type, $F(2, 116) = 54.20$, $MSE = 421.21$, $p<.001$, can be accounted for by the low strength of conclusion ratings by participants on problems that were Inconsistent ($p<.001$ when compared to Consistent and Neutral problems). That is, the problems which contain conclusions which are inconsistent with a person's beliefs and elicit more LLN reasoning are perceived as weaker arguments than the ones that contain conclusions which are consistent or neutral to a person's belief system.

The ANOVA also yielded a significant interaction between condition and problem type ($F(2, 116) = 3.28$, $MSE = 25.37$, $p<.05$). Follow-up analyses revealed that after training, participants rated the arguments as stronger on neutral items (at $F(1, 58) = 6.95$, $MSE = 60.00$, $p<.05$). Training did not effect the ratings on consistent and inconsistent items ($F(1, 58) = 0.35$, $MSE = 5.40$, $p>.1$ and at $F(1, 58) = 0.5$, $MSE = 6.02$, $p>.1$ respectively).

A 2 (Condition) x 3 (Problem Type) mixed ANOVA was performed on the ratings on the persuasiveness scale. No effect of training was found ($F(1, 58) = 0.45$, $MSE =6.05$, $p>.1$). Participants rated the Inconsistent problems as less convincing than Consistent or Neutral problems ($F(2,116) = 53.69$, $MSE = 431.82$, $p<.001$; a follow-up analysis found differences between Inconsistent and Consistent and Neutral to be significant at $p<.001$ for both). The interaction failed to attain significance ($F(2, 116) = 2.20$, $MSE = 17.72$, $p>.1$).

4.4.2 Experiment Evaluation Reasoning

Responses on the experiment evaluation problems were scored according to the 3-point coding system as described in the Method section. As with the LLN problems, scores were then collapsed across the three problem types, Neutral, Consistent and Inconsistent.

It may be seen in Table 4.3 that there is no apparent difference in quality of experiment evaluation responses dependent on whether the participants received training or not, but it does appear as in LLN reasoning, that more sophisticated reasoning is used on problems involving belief-Inconsistent conclusions. A 2 (Condition) x 2 (Problem type) ANOVA was performed with Condition as a between-subjects factor and Problem Type as a Within-subjects factor. There was no effect of training on experiment evaluation reasoning ($F(1, 58) = 0.51$, $MSE = 1.80$, $p>.1$) but there was a significant difference between the problem types in levels of sophistication of reasoning ($F(2, 116) = 68.12$, $MSE = 98.15$, $p<0.001$) and an LSD follow-up analysis shows that the three problem types elicit significantly different levels of reasoning. Belief-Inconsistent problems elicit the most sophisticated reasoning style, followed by Belief-Neutral problems and finally Belief-Consistent problems (all $p<0.001$). In other words, as with LLN reasoning, problems with conclusions that are inconsistent with ones beliefs will elicit a deeper reasoning strategy than problems with either consistent or neutral conclusions. No interaction was obtained between Condition and Problem Type ($F(2, 116) = 2.74$, $MSE = 3.95$, $p>.05$) .

	Neutral	Consistent	Inconsistent
control	2.93 (1.68)	1.13 (1.07)	3.53 (1.87)
training	2.17 (1.18)	1.06 (1.05)	3.76 (1.69)

Table 4.3 Mean scores for Argument Evaluation Reasoning on each type of problem under each condition (N=60: Standard deviations shown in brackets)

4.4.2.1 Inappropriate use of LLN reasoning

Also of interest was whether participants in the Training condition would use law of large numbers reasoning inappropriately. The experiment evaluation problems were not designed to elicit LLN reasoning therefore any use of this reasoning strategy would be incorrect.

Fong et al. found no overuse of the LLN principle in their study. They included 'false alarm' problems to explore whether their training promoted the overuse of LLN on problems where it was not appropriate. Fong et al. concluded that their training did not lead to widespread overuse of the law of large numbers and indeed trained participants are able to use the law discriminately. In this experiment inappropriate use was scored on items when the participant stated that the sample size was too small. Overall, for participants who received training, inappropriate use of LLN was noted on only 9% of the problems totally. However nearly 50% of the participants used it on at least one problem erroneously compared to the control group who did not refer to sample size on a single occasion.

4.4.2.2 'Conclusion strength' and 'Experiment Validity' Scales

Following each problem, participants were asked to indicate on two 9-point scales how strong they thought the researcher's conclusion was (1= extremely weak, 9= extremely strong), and how valid they thought the experiment was (1= extremely invalid, 9= extremely valid). Total scores were collapsed across problem type, Neutral, Consistent and Inconsistent, which resulted in two rating scales, the 'conclusion strength' and 'experiment validity' scales respectively. Scores on each rating scale ranged from 3 to 27 for each problem type.

It can be seen in Table 4.4 that the evidence involving conclusions which are inconsistent with belief are judged as weaker that either belief-neutral or belief-consistent problems. A 2 (Condition) x 3 (Problem type) ANOVA revealed that there was a significant effect of Problem Type ($F(2,116) = 44.32$, $MSE = 395.40$, $p<0.001$) but not Condition ($F(1, 58) = 2.80$, $MSE = 115.2$, $p>.1$). The interaction between condition and problem type also failed to attain significance (at $F(2, 116) = 1.73$, $MSE = 15.4$, $p>.1$). An LSD follow-up analysis revealed that all three problem types elicited significantly different mean rating scores from each other (p<0.001), with belief-consistent problems having the 'strongest evaluation' ratings, belief-neutral the next and belief-inconsistent problems having the 'weakest evaluation' ratings. This is consistent with the findings on the LLN problems.

Table 4.4 also displays the mean ratings on the 'experiment validity' scale for each problem type. Again the lowest scores appear to be for the Belief-inconsistent problems and the 2 (Condition) x 3 (Problem type) ANOVA performed showed that there was a main effect of Problem Type and Condition ($F(2,116) = 47.23$, $MSE = 284.30$, $p<.001$ and $F(1,58) = 11.55$, $MSE = 341.70$, $p<.01$ respectively), with 'validity' scores being higher for participants in the Training condition. A follow-up analysis on the effect of Problem Type found significant differences between scores on all three ($p<.001$ for all).

Scale	Problem Type	Control		Training	
		Mean	St.dev.	Mean	St.dev.
Conclusion Strength	Neutral	10.50	4.70	13.27	4.42
	Consistent	14.50	4.22	15.60	4.47
	Inconsistent	9.50	4.26	10.43	4.54
Experiment Validity	Neutral	9.87	4.12	13.90	3.49
	Consistent	12.87	3.88	15.57	3.43
	Inconsistent	9.10	3.57	10.63	3.82

Table 4.4 Mean ratings on the 'Conclusion strength' and 'Experiment validity' scales for each problem type under each condition.

An interaction between Condition and Problem Type was also found, $F(2,116) = 3.90$, $MSE = 23.50$, $p<.05$, see Figure 4.2. This shows that training on LLN principle significantly increased validity ratings for Neutral and belief-consistent problems (at $F(1, 58) = 16.77$, $MSE = 244.02$, $p<.001$; and $F(1, 58) = 8.14$, $MSE = 109.35$, $p<.01$ respectively) but not for belief-inconsistent problems ($F(1, 58) = 2.58$, $MSE = 35.27$, $p>.1$). In other words, participants who received training rated belief-neutral and belief-consistent problems as even more valid, but validity ratings were not much higher for belief-inconsistent problems after training than for the Control group. This suggests that participants rated the validity of the experiments as generally higher after training, perhaps because all of the problems involved large samples which they then judged as more important after training or because the training led the participants to focus on the sample size, for instance, larger sample therefore more valid.

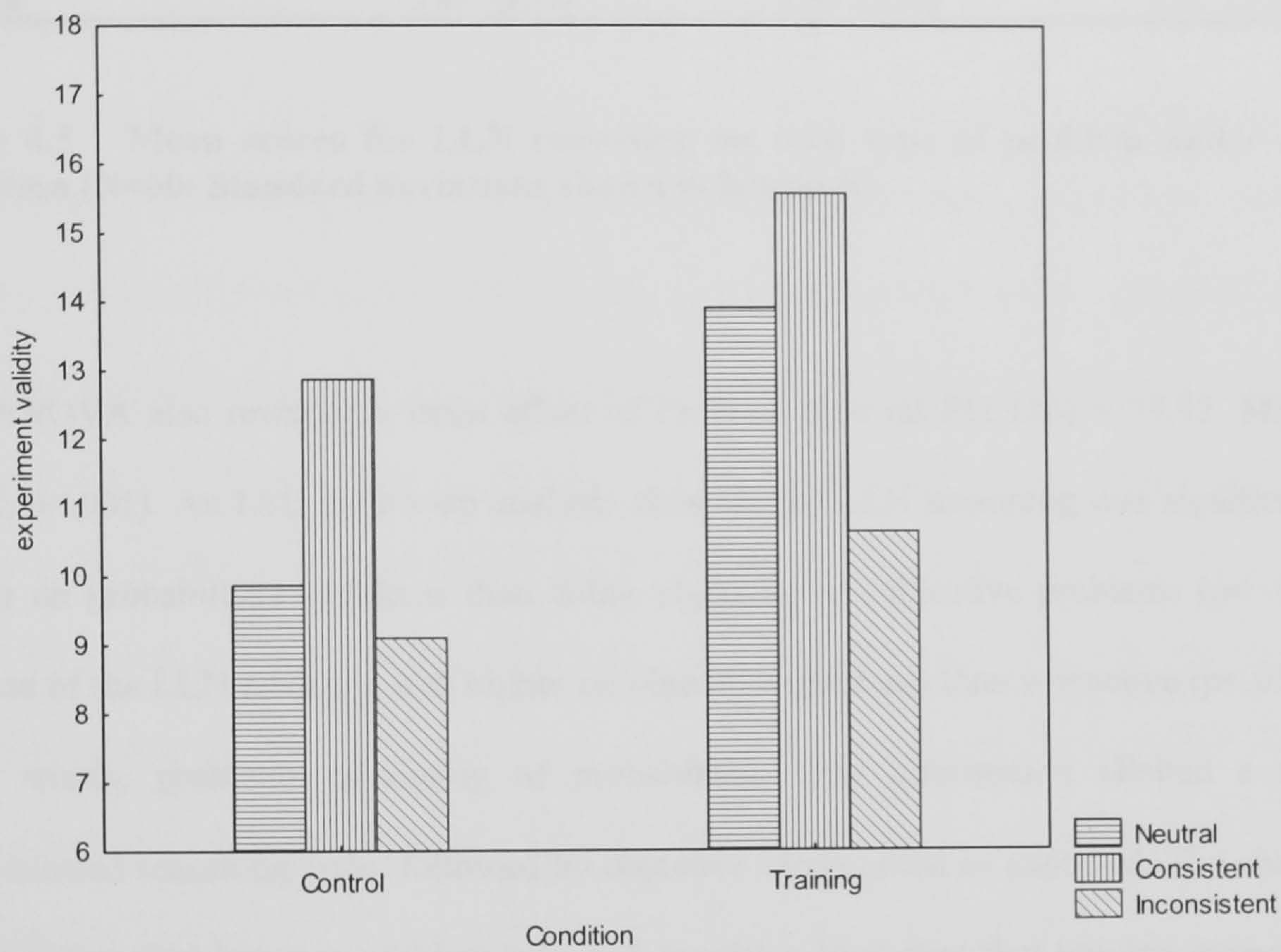

Figure 4.2 Interaction between Condition and Problem Type on the 'Experiment Validity' scale.

4.4.3 Law of Large Numbers reasoning – Fong type problems

Participants' responses on the nine problems were scored using the 3-point system described in the Method section. Scores were then collapsed across the three problem types, probabilistic, objective and subjective. Unlike the other LLN problems, these problems did not consist of any manipulation of belief.

Table 4.5 displays the mean scores for LLN reasoning on each type of problem under each condition. Probabilistic reasoning elicited the most LLN reasoning in the control condition as expected. After training on the law of large numbers, LLN reasoning was improved on all the tasks. A 2 (Condition) x 3 (Problem type) ANOVA with repeated measures on the second factor revealed a main effect of condition (at $F(1,58) = 26.40$, $MSE = 74.76$, $p<.001$).

	Objective	Probabilistic	Subjective
Control	0.77 (0.93)	2.23 (1.50)	0.57 (0.90)
Training	2.40 (1.71)	3.50 (1.48)	1.53 (1.36)

Table 4.5 Mean scores for LLN reasoning on each type of problem under each condition (N=60: Standard deviations shown in brackets)

The ANOVA also revealed a main effect of Problem type (at $F(2,116) = 39.92$, $MSE = 52.32$, $p<.001$). An LSD follow-up analysis showed that LLN reasoning was significantly higher on probabilistic problems than either objective or subjective problems ($p< .001$), and use of the LLN principle was higher on objective problems than subjective ($p<.05$). In other words, problems consisting of probabilistic type information elicited a more sophisticated reasoning style, followed by objective information as expected. The absence of any interaction between problem type and condition illustrates that training impacts on all three types of problem equally ($F(2, 116) = 1.28$, $MSE = 1.67$, $p>.1$).

4.5 Discussion – Experiment 2

In summary, after training on the law of large numbers, use of statistical principles was increased on all the problems designed to elicit belief-based responses and Fong et al.'s everyday reasoning problems involving probabilistic, objective and subjective content, even after a one-week delay. Klaczynski's findings were replicated in this experiment. Arguments or experiments involving belief-inconsistent conclusions elicited more sophisticated reasoning strategies than either belief-neutral or belief-consistent conclusions. On law of large numbers problems, participants utilised the LLN principle and on experiment evaluation problems participants identified more flaws in the hypothetical research. Training had no effect on experiment evaluation reasoning.

The effects of training on the everyday reasoning problems designed to elicit belief-based responses were surprising. Participants were able to utilise the LLN principle on all problems, whether they were neutral, consistent or inconsistent with their beliefs. It was predicted that the effects of bias may be reduced as in the syllogistic reasoning literature (Evans et al., 1994). In accord with the belief bias literature, Stanovich (personal communication, October 2004) proposed that training in rule-based strategies would attenuate but not eliminate biases. In dual process terms, instruction would increase System 2 function which would inhibit System 1 responses. In contrast Klaczynski suggested (personal communication, November 2004) that the training may not even transfer to problems involving belief-laden content at all. He argues that the two systems are independent and instruction may increase LLN reasoning but would have no impact on the belief-based responses at all. However the results reported in this experiment illustrate the elimination of bias after training on the concept of the law of large numbers.

The influence of belief itself was still present in the argument ratings. Examination of the ratings after each problem demonstrated that training did not eliminate the influence of beliefs on ratings related to the persuasiveness of the argument and the strength of the conclusion based on the evidence. Participants still rated the evidence which was inconsistent with beliefs as weaker or less convincing than both consistent and neutral evidence. This illustrates a dissociation between analytic and belief-based influences between the ratings and the written justifications, as demonstrated by the fact that individuals can utilise the LLN principle on belief-consistent and inconsistent problems after training, hence utilising System 2 processes. However, evidence for the influence of beliefs is still present in the ratings of the arguments.

There were no effects of training on the experiment evaluation problems which was as expected. However, more sophisticated reasoning strategies were used on the problems involving belief-inconsistent conclusions. In other words, participants were more likely to search for and identify confounds in the experimental design which would invalidate the studies when the evidence was contrary to beliefs. Again, when the evidence was consistent or neutral to a person's beliefs then the conclusion was much more likely to be accepted with little or no argument.

Participants rated belief-inconsistent conclusions on the experiment evaluation problems as weaker and the evidence less valid than belief-consistent or neutral conclusions and evidence. These findings again replicate those of Klaczynski and associates. However after LLN training validity ratings on belief-consistent and neutral problems were increased. This appears to be an overgeneralization of the LLN training. It is possible that the training redirected participants' attention to the wrong components of these problems. They only paid cursory attention to these problems in the first place because they involved conclusions that were consistent or neutral to their beliefs, they recognised that the sample sizes were pretty large; hence they concluded that the experiments must be more valid.

From the findings reported here and those reported in the literature, people do appear to have use of an intuitive general, domain independent rule that states 'the larger the sample the better'. Without training in the inferential rule system, people utilise the implicit rule automatically on problems designed to elicit the rule e.g. probabilistic problems and belief-inconsistent problems. Given training on the rule system, participants are then able to use explicit principles on a range of different problems outside the domain of training.

Experiment 3 had two aims. Firstly to replicate these findings. Secondly, if explicit training on the concept of the law of large numbers could increase rational System 2 processing.

was there a different explicit instructional approach that could result in inhibition of System 1 and therefore less biased responding?

4.6 Introduction and Rationale for Experiment 3

In the previous experiment it was demonstrated that training in statistical principles led to more sophisticated reasoning strategies on a variety of everyday reasoning problems. Consequently the effect of bias in participants' responses was also eliminated. The aim of this experiment was to replicate the main findings of Experiment 2 using the same methodology, and to extend on it by including a second instructional condition. If System 2 processing can be improved by formal training in the law of large numbers which indirectly leads to the elimination of bias, can it be improved by an instruction technique designed to encourage participants to disassociate from their beliefs?

Both Klaczynski and Gordon (1986) and Stanovich (1999) argue that motivational style is a predictor of decontextualised reasoning. Klaczynski, Gordon, and Fauth (1997) proposed that decontextualised reasoning is a metacognitive competence independent of intelligence. They found that bias was predicted by information processing style. Individuals with a more rational thinking style were less biased in their reasoning. These individuals were able to decontextualise the key elements from the content of the problem. Stanovich (1999; Stanovich & West, 1997) argued that both cognitive ability and thinking dispositions were predictors of decontextualised reasoning; ability at the computational level of analysis and thinking dispositions at the intentional level of analysis. It was proposed that individuals with higher ability were able to flexibly contextualise and decontextualise as necessary.

It is proposed that in contrast to ability which is fixed, thinking styles are malleable therefore more teachable. In which case using instruction procedures to motivate people to improve performance should have some effect. Houdé and Moutier (1999) proposed that motivational styles were an important factor in training. They found that it was individuals with a more field-independent cognitive style that were more susceptible to inhibition training on the selection task. These individuals were more able to perceive an element separate from its geometric content and to adopt an analytic attitude in problem solving.

Klaczynski and Gordon (1986) investigated whether the motivation to construct 'correct' solutions could improve performance and/or decrease biases. They gave participants 'accuracy motivation' instructions. These informed participants that if they were to give confusing, thoughtless or inaccurate responses, they would be contacted the following week. They would then be required to meet with the experimenters to justify and clarify their responses. Overall performance was increased after these instructions. Bias, however was not reduced. Thus it appeared that the motivation to defend beliefs was stronger than the motivation to be accurate.

One of the aims of Experiment 3 was to investigate whether an instructional manipulation designed to foster cognitive decontextualisation would decrease biased responding on the everyday reasoning tasks. The instructions were designed using the Complex Instructions in the belief bias literature as a model (Evans et al. (1994) see Example 2.2. Chapter 2), with the emphasis placed on the assumption of truth in the problems and disassociation from what is believed to be true. If cognitive motivation is an important factor in decontextualised reasoning, then instructions to disengage from beliefs should have an effect. However, if belief-based responses are intuitive, automatic responses, participants may not be able to inhibit them.

It was predicted that the findings of Experiment 2 would be replicated. Performance under the Control and the Training conditions would be consistent with performance in the previous experiment. For participants in the Complex Instruction condition, bias in responding will be reduced on the LLN and Experiment Evaluation problems as the instructions aim to foster decontextualised reasoning strategies.

4.7 Method

Design

This experiment was designed to both replicate the findings of Experiment 2 and extend on the findings by including another condition. It had a between subjects design involving three conditions, a Control group that just received minimal instructions and a full training group (both as in Experiment 2), and a Complex instruction group which had not been used previously. The instructions (see below), were designed using the Complex Instructions in the belief bias literature as a model (Evans, Newstead, Allen, & Pollard, 1994). These instructions served to reduce, though not eliminate, belief bias in syllogistic reasoning. All participants received a test booklet consisting of law of large numbers problems and experiment evaluation problems as used in Experiment 2.

Participants

The participants were 90 undergraduates (81 females, 9 males) from the University of Plymouth, taking part for course credit, randomly allocated to each group resulting in 30 per experimental condition.

Materials

Instructions:

-Participants in the Control group were presented with the instructions described in Experiment 2 followed by the test booklet.

-Participants in the Full training condition were given the full written instructions followed by the demonstration then the example problems as in Experiment 2, followed by the test booklet.

Complex Instruction Condition

The Complex instructions had not been used previously and were designed to observe whether an instructional manipulation may elicit the use of more sophisticated reasoning strategies (see Table 4.6 for the full script).

The Complex instructions script was partly designed using the instructions from the belief bias in syllogistic reasoning research conducted by Evans, Newstead, Allen, and Pollard (1994, Experiment 3) as a framework. Evans et al. found that their Complex instructions decreased bias (though didn't eradicate it altogether) in syllogistic reasoning. Their instructions involved a short description of the syllogistic reasoning task, emphasised the use of deductive logic and asked participants to assume that the information given in the syllogisms was true. The instructions used for this study gave a short description of the types of problems in the test booklet but also emphasised to participants the assumption that all the information given is true and that responses must be based on this and not on what they believe to be true, see Table 4.6 below for full script. It was hoped that these instructions would elicit the use of analytic reasoning (System 2 responding), therefore inhibiting pre-conscious belief-based responding (System 1 responding).

Complex Instructions

This experiment is designed to find out how people solve and evaluate problems involving real-life content.

In this booklet that you have been given there are 2 different types of problems – 1 type consists of reasoning problems that you may come across yourself in everyday life. You will be presented with a short scenario to read and then you will be asked to indicate on a scale, based on the evidence used, how weak or strong you think the conclusion is and on another scale how convinced you are by the argument. You will then be asked to comment on the thinking in the argument in no more than 2 or 3 sentences, explaining why you think the conclusion is good or bad, weak or strong and convincing or not convincing.

Table 4.6 Continued:

Since these problems require rational analysis, you should base your answers on the content of the scenarios, not on what you believe to be true. If, and only if, you judge that a conclusion follows from the information given should you rate it as strong and convincing. The conclusion given may not always be the same as what you believe.

The second type of problem involves summaries of actual research which you will be asked to comment on. After reading a summary you will be asked to rate the strength of the researchers conclusion and the validity of the experiment on two scales. You are then required to explain why you think the researcher's conclusion was weak or strong and why you think the experiment was valid or invalid.

Again, you are asked to analyse the research based on the content of the summaries, not on what you believe to be true. If, and only if, you judge that the researchers conclusion follows from the summary presented should you rate it as strong and valid. The conclusion given may not always be the same as what you believe.

Please take your time and be certain that you have made the logically correct decision before stating it. If you have any questions, please ask them now, as the experimenter cannot answer any questions once you have begun the experiment. Please keep these instructions in front of you in case you need to refer to them later on.

YOU ARE REMINDED THAT YOU MUST BASE ALL YOUR RESPONSES ON THE INFORMATION GIVEN IN THE ARGUMENT SCENARIOS AND RESEARCH SUMMARIES – AND THIS INFORMATION ONLY. YOU MUST ASSUME THAT ALL THE INFORMATION WHICH YOU ARE GIVEN IS TRUE – THIS IS VERY IMPORTANT. DO NOT BASE YOUR RESPONSES ON WHAT YOU BELIEVE TO BE TRUE.

Please do not turn back and forth from one problem to another once you have started. You may now continue to work through the booklet.

Table 4.6 Complex Instructions used in Experiment 3

Test Materials:

Eighteen problems involving a belief manipulation were used in total, nine Law of Large Numbers and nine Experimental Evaluation problems as used in Experiment 2.

Coding

The same 3 point coding systems used in Experiment 2 were utilised in this experiment. See Experiment 2's Method section for details.

For law of large numbers reasoning, total agreement was obtained between the principal researcher and the independent coder, blind to condition and problem type, on 94% of the problems. For experiment evaluation reasoning, total agreement was obtained on 96% of the problems. Any disagreements were discussed and agreed upon.

Procedure

Participants were randomly allocated to experimental conditions in groups of 2 to 6. The Control group were given their instructions and test booklet and were given an hour to complete the tasks. The Complex instruction group were given their instruction sheet and asked to take their time in reading it. They were then given the test booklet but reminded to refer to the instructions periodically. The Full Training group were asked to read the full instructions and were then given the demonstration by the experimenter. Following this they were asked to complete the example problems. On completion they were asked to continue through the test booklet. No delay was included between training and testing as in Experiment 2 due to the large effects even after a delay. Without a delay the instruction and training conditions could be directly compared. Participants in the Full Training condition were given 1 1/2 hours to complete the whole experiment.

On completion of the experiment, participants were thanked and debriefed.

4.8 Results

4.8.1 Law of Large Numbers reasoning

Items were scored as described for Experiment 2. Means and standard deviations are illustrated in Table 4.7. Note that the scores after training are higher in this experiment than in the previous experiment. This is probably due to the lack of delay between training and testing. Problems consisting of conclusions that were inconsistent with a person's beliefs yield the highest scores, which is consistent with the previous findings that individuals utilise a more sophisticated reasoning strategy on these problems. It is also evident that higher LLN scores are attained after training on the concept of statistical reasoning as also found previously. A 3 (Condition) x 3 (Problem type) ANOVA conducted on the data resulted in a significant effect of condition ($F(2, 87) = 35.90$, $MSE = 159.60$, $p<.001$). An LSD follow-up confirmed that participants in the Training condition attained higher LLN scores than either the Control group or the Instruction group ($p<.001$ for each). A significant effect of Problem Type was also found ($F(2, 174) = 13.09$, $MSE = 18.98$, $p< .001$) with Inconsistent problems yielding higher LLN scores than either Consistent problems or Neutral problems ($p<.01$ and .001 respectively), again as expected. Higher scores were obtained on the Consistent problems than the Neutral ones ($p<.05$).

	Neutral	Consistent	Inconsistent
Control	1.53 (1.50)	1.87 (1.67)	3.00 (1.64)
Instruction	1.83 (1.46)	2.23 (1.69)	2.93 (1.63)
Training	4.37 (1.45)	4.70 (1.58)	4.53 (1.41)

Table 4.7 Mean LLN score on the three problem types under each condition (N=30 per condition. Standard deviations shown in brackets).

The significant interaction obtained between Condition and Problem Type is shown in Figure 4.3 ($F(4, 174) = 3.06$, $MSE = 4.44$, $p<.05$). The graph serves to illustrate that LLN reasoning is elicited when participants are given problems that involve conclusions that are inconsistent with their beliefs under the Control and the Instruction conditions, but after training participants utilise statistical principles to the same extent on each of the problem types. Under the Control condition, scores on Inconsistent problems were significantly higher than scores on Consistent and Neutral problems ($F(1, 87) = 11.98$, $MSE = 19.27$, $p<.001$; $F(1, 87) = 23.96$, $MSE = 32.27$, $p<.001$ respectively). In other words, when presented with problems or arguments that are inconsistent with prior beliefs, participants utilise a more sophisticated reasoning style to argue with, whereas if the information is consistent or neutral to beliefs then it is much less likely to be evoked.

Under the instruction condition there are no effects of instruction on use of statistical reasoning and planned comparisons reveal that any increase was not significant (F(1, 87) = 0.62, MSE = 1.35, p>.4; F(1, 87) = 0.74, MSE = 2.02, p>.3 belief-neutral and belief-consistent respectively).

After training on the LLN principle, scores on all the problem types were greatly increased. There are no differences in levels of sophistication of responding on the different problem types (F(1, 87) = 1.19, MSE = 1.67, p>.2 between neutral and consistent problems; F(1, 87) = 0.26, MSE = 0.42, p>.6 between consistent and inconsistent problems; F(1, 87) = 0.31, MSE = 0.42, p>.5 between neutral and inconsistent problems). Training in statistical principles leads to more sophisticated reasoning strategies on all three problem types.

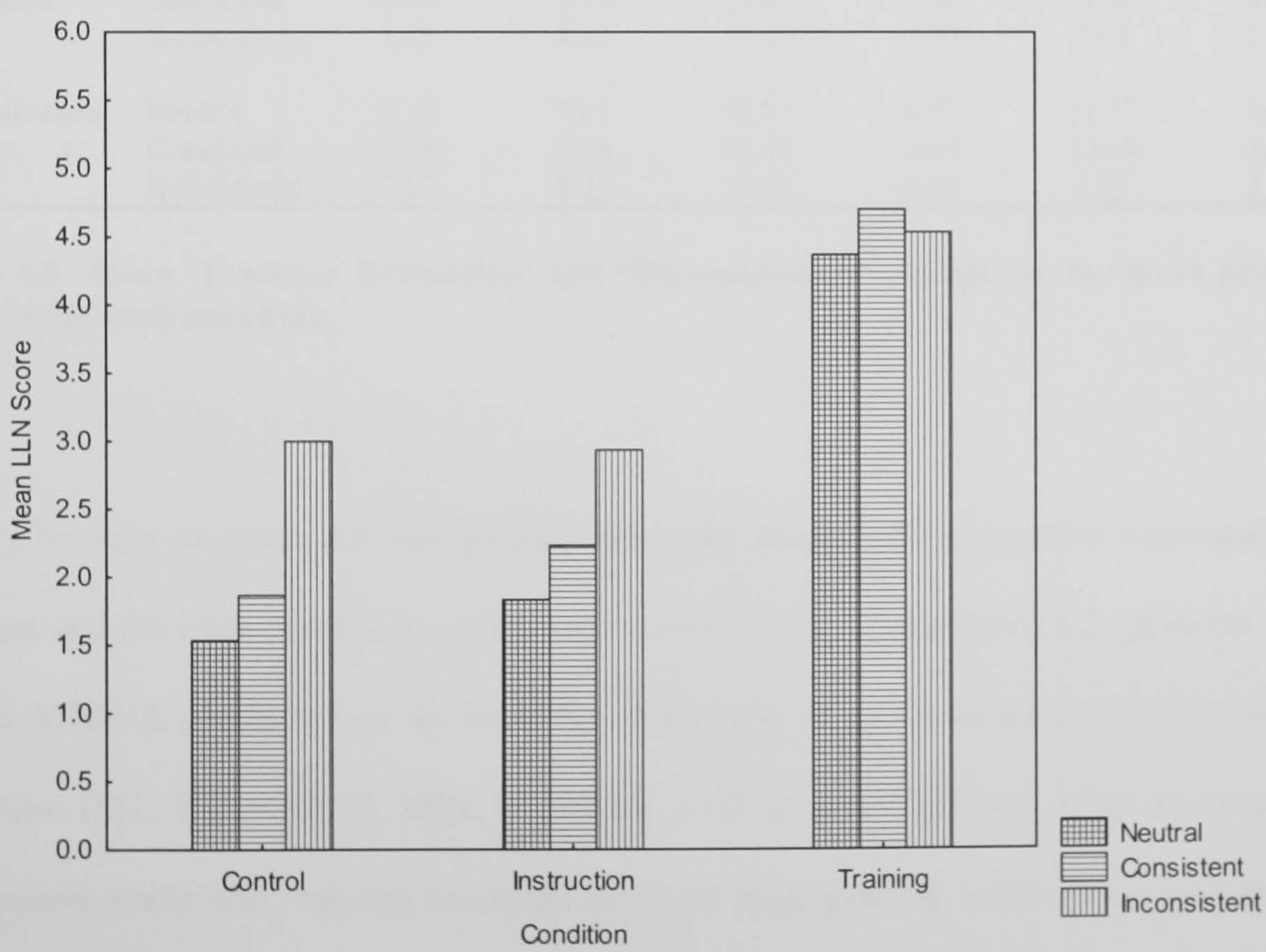

Figure 4.3. Interaction between Condition and Problem Type on LLN reasoning.

4.8.1.1 'Evidence Evaluation' and 'Persuasiveness Scales'

Following each problem, as in Experiment 2, participants indicated on two 9-point scales based on the evidence presented, how weak or strong they thought the conclusions were, (1=extremely weak, 9= extremely strong), and how convinced they were by the argument (1=Extremely unconvinced, 9=Extremely convinced). Total scores were collapsed across problem type, Neutral, Consistent and Inconsistent, which resulted in the two rating scales, the 'evidence evaluation' and 'persuasiveness' scales respectively. Scores on each rating scale ranged from 3 to 27 for each problem type, see Table 4.8 for the means on both scales.

Scale	Problem Type	Control		Instruction		Training	
		Mean	St.d.	Mean	St.d.	Mean	St.d.
Evidence Evaluation	Neutral	12.63	3.37	14.90	3.23	12.33	3.60
	Consistent	15.03	5.17	16.67	3.59	11.80	4.00
	Inconsistent	8.93	4.23	11.40	4.30	7.03	2.76
Persuasiveness	Neutral	11.47	3.21	13.57	3.57	11.77	3.87
	Consistent	14.20	5.06	15.70	3.05	12.00	4.68
	Inconsistent	7.57	3.27	10.40	3.55	7.23	3.10

Table 4.8 Mean 'Evidence Evaluation' and 'Persuasiveness' ratings for the three problem types under each condition

It may be seen in Table 4.8 that participants rated the belief-inconsistent arguments the weakest and the least convincing under each condition. A 3 (condition) x 3 (problem type) mixed ANOVA conducted on the evidence evaluation ratings resulted in a main effect of condition ($F(2, 87) = 13.63$, $MSE = 348.83$, $p<.001$). An LSD follow-up showed that participants under the Training condition rated the arguments as weaker than participants in either the Control or Instruction condition ($p<.05$ and .001 respectively) and the participants in the Instruction group rated the arguments as stronger than participants in the Control group ($p<.01$). It appears that the LLN training may alert participants to the weakness of the arguments. However asking participants to assume the information is true in the instruction condition may lead participants to have less doubt in the evidence leading them to rate the arguments as strong.

The ANOVA also resulted in a main effect of Problem Type ($F(2, 174) = 74.53$, $MSE = 716.23$, $p<.001$) with participants rating problems involving belief-consistent information as stronger than either inconsistent or neutral problems ($p< .001$ and .01 respectively) and problems involving neutral content as stronger than problems involving inconsistent content. No significant interaction between Condition and Problem Type was obtained ($F(4, 174) = 1.98$, $MSE = 19.05$, $p>.05$).

A 3 (Condition) x 3 (Problem type) mixed ANOVA performed on responses on the persuasiveness scale, resulted in a main effect of Condition ($F(2, 87) = 8.32$, $MSE = 202.48$, $p<.001$) and an LSD follow-up test revealed there was no significant difference between the Training group and the Control group ($p=.31$). Participants in the Instruction group rated the arguments as more convincing than participants in the Control group and the Training group ($p<.01$ and $p<.001$ respectively). This is consistent with the findings on the evidence evaluation scale where participants rate the arguments as stronger after instruction to dissociate from beliefs.

The ANOVA also yielded a significant main effect of problem type ($F(2, 174) = 80.29$, $MSE = 732.43$, $p<.001$) and the follow-up analysis revealed that Inconsistent problems were rated as significantly less convincing than Consistent and neutral problems ($p<.001$ for both). Neutral problems were rated as significantly less convincing than Consistent problems ($p< .001$). No significant interaction between Condition and Problem type was obtained ($F(4, 174) = 1.70$, $MSE = 15.48$, $p>.1$).

There are two key findings here in relation to the rating scales. Firstly when asked to evaluate the strength of the conclusion based on the evidence presented, participants are strongly influenced by their beliefs, even when they are given training that provides them with the skills necessary to objectively evaluate the evidence. However they are not influenced by beliefs when they are subsequently asked to provide a written evaluation of the argument, as shown by the earlier analysis of levels of LLN reasoning after training. Therefore the training is eliminating the effects of belief on how they justify their evaluations, but it is not influencing their ratings of the evidence. Secondly, when participants are under instruction to dissociate from beliefs, participants rate the evidence strength and persuasiveness as higher than either of the other conditions. It is as if the instructions to 'assume the information is true' removes the reason for evaluating the evidence strength or persuasiveness as low. Participants are assuming that the information is true therefore it is strong, rather than that the evidence itself is inaccurate. However the relative influence of belief is not changed after instruction. We will return to both of these findings in the general discussion that follows.

4.8.2 Experiment Evaluation Reasoning

Items were coded using the 3-point coding system as described for the previous experiment. Scores were collapsed across problem type and the mean experiment evaluation scores under each condition may be observed in Table 4.9.

	Neutral		Consistent		Inconsistent	
Control	2.83	(1.37)	1.77	(1.45)	4.07	(1.20)
Instruction	2.97	(1.85)	1.73	(1.46)	3.53	(1.98)
Training	2.43	(1.38)	1.50	(1.61)	3.20	(1.86

Table 4.9 Mean Experiment Evaluation Scores for each problem type under each condition (N=30 per condition. Standard deviations shown in brackets).

A 3 (Condition) x 3 (Problem Type) mixed ANOVA showed there was no main effect of condition ($F(2, 87) = 1.32$, $MSE = 6.25$, $p>.1$). The ANOVA revealed a main effect of Problem Type ($F(2, 174) = 58.74$, $MSE = 84.47$, $p< .001$). A follow-up analysis shows problems involving belief-inconsistent conclusions elicit significantly higher scores than both problems consisting of belief-consistent or belief-neutral conclusions and belief-consistent problems yield significantly lower scores than the belief-neutral problems (all $p<.001$). These results are all in the expected direction and replicate Experiment 2. There was no interaction between Condition and Problem Type ($F(4, 174) = 0.85$, $MSE = 1.21$, $p>.1$).

These findings show that when participants are given problems that involve conclusions that are inconsistent with their beliefs, they are likely to work harder at finding the flaws in the evidence than when given problems involving conclusions that are consistent with what they believe. If the conclusion is consistent with their beliefs, then the participants will readily accept the evidence. Training on statistical concepts or instruction to disassociate from belief does not have any effect on responses.

4.8.2.1 Overuse of LLN reasoning

Fong et al. concluded that training did not lead to widespread overuse of the law of large numbers and indeed found that participants were sophisticated in avoiding the improper use of the law of large numbers. Participants overuse the LLN principle after training by stating that the sample size is too small therefore the experiment is invalid. On analysis of individual responses in this experiment, Inappropriate use of statistical principles was utilised on nearly 11% of items overall and 60% of participants stated that the sample was too small on at least one item. Participants in the Control condition did not utilise law of large numbers reasoning on these items. This would lead to the conclusion then that there is some overuse of the law of large numbers on experiment evaluation problems.

4.8.2.2 'Conclusion strength' and 'Experiment validity' scales

Following each problem, participants were asked to indicate on two 9-point scales how strong they thought the researcher's conclusion was (1 = extremely weak, 9 = extremely valid), and how valid they thought the experiment was (1 = extremely invalid, 9 = extremely valid). Total scores were collapsed across problem type, Typical, Untypical and Neutral. This resulted in two rating scales, the 'conclusion strength' and 'experiment validity' scales respectively (see Table 4.10). Scores were collapsed across problem type and ranged from 3 to 27 for each.

It can be seen in Table 4.10 that problems involving conclusions that are belief-consistent are rated as a lot stronger that problems involving conclusions that are belief-inconsistent. Participants in the Control condition appear to rate the experiments as slightly weaker than those in the Training and Instruction conditions. A (3 (Condition) x 3 (Problem Type) ANOVA resulted in a main effect of Problem Type ($F(2, 174) = 79.70$, $MSE = 690.78$, $p<.001$), and the follow-up analysis revealed that belief-consistent problems were perceived as stronger than both belief-inconsistent and neutral problems ($p<.001$ for each). Belief-neutral problems were viewed as stronger than Belief-inconsistent ones ($p< .01$). In other words, when presented with problems involving conclusions that are inconsistent with their prior beliefs, participants will rate them as involving weaker evidence than the other problems involving conclusions which are consistent with or neutral to their beliefs. A main effect of Condition was also obtained ($F(2, 87) = 3.19$, $MSE = 107.7$, $p<.05$) and the follow-up revealed that participants in the Control condition rated the evidence as slightly weaker than participants in the other two Training and Instruction conditions ($p< .05$ for both). No interaction between training and problem type was obtained (at $F(4, 174) = 1.86$, $MSE = 16.10$, $p>.1$).

For the 'Experiment validity' scale, participants rate belief-inconsistent problems as less valid than belief-consistent problems and participants in the Control group rate all the problems as slightly less valid than participants in the other two conditions. The 3 (Condition) x 3 (Problem type) mixed ANOVA yielded a main effect of training condition (at $F(2, 87) = 4.76$, $MSE = 136.58$, $p<.05$). An LSD follow-up found that participants under the Training condition rated the problems as more valid than participants in the control group ($p<.01$).

Scale	Problem Type	Control		Instruction		Training	
		Mean	St.d.	Mean	St.d.	Mean	St.d.
Conclusion	Neutral	10.27	4.26	12.37	3.51	13.57	4.20
Strength	Consistent	15.53	4.56	16.67	3.87	16.30	3.79
	Inconsistent	9.53	4.04	11.97	4.78	11.17	3.98
Experiment	Neutral	10.13	3.54	11.77	3.75	13.53	3.73
Validity	Consistent	13.97	4.31	15.00	3.37	16.33	3.73
	Inconsistent	9.30	4.17	9.50	3.59	10.87	4.49

Table 4.10 Mean scores on the 'Conclusion strength' and 'Experiment validity' Scales

The ANOVA also yielded a main effect of problem type ($F(2, 174) = 76.93$, $MSE = 625.01$, $p<.001$). As expected, participants rate the evidence on Inconsistent problems as less valid than either Consistent or Neutral problems ($p< .001$ for each). In other words, if individuals disagree with the conclusion they rate the evidence as less valid than if the conclusion is consistent or neutral with their beliefs. No interaction between condition and problem type was obtained (at $F(4, 174) = 0.87$, $MSE = 7.11$, $p>.4$)

4.9 Discussion – Experiment 3

The findings of Experiment 2 were replicated in this experiment. Training on the concept of the law of large numbers led to use of the statistical principles on problems designed to elicit belief-based responses resulting in the elimination of bias. The instructional manipulation to motivate participants to dissociate from beliefs had no effect on participants' responses on any of the problem types compared to the control group.

Once again, this experiment has demonstrated that training in statistical principles increases the sophistication of reasoning responses on problems involving belief-consistent, belief-inconsistent and neutral content. The influence of belief was eliminated in terms of the effect on sophistication of LLN reasoning demonstrated in responses. However evidence from the rating scales illustrates that the influence of belief was still present. Belief-consistent arguments were still viewed as stronger and more convincing than belief-inconsistent arguments, again showing a dissociation between the two response measures. These findings will be discussed further in the general discussion.

The instructional procedure utilised in this experiment failed to improve reasoning on either LLN problems or experiment evaluation problems. Evans, Newstead, Allen, and Pollard (1994) found a reduction in belief-based responding after an instruction procedure which emphasised the structure of the syllogism, explained the logical meaning of 'SOME' and also emphasised the need to base responses only on the information given. In this experiment, there was no explanation of the law of large numbers, the rule to utilise on the LLN problems, or any information about experimental confounds built into the instruction set. Taking the findings from the LLN training, it appears that people need to be given the rule to follow, and simply instructing them to ignore beliefs is not sufficient. According to Sloman (personal communication, September 2004) people need to be cued to use a rule, so bias will be reduced only to the extent that the cue is effective. In the instructional condition there were no cues for either type of problem.

It is clear here and in Experiment 2 that law of large numbers reasoning is utilised on belief-inconsistent problems in the absence of training or a cue. The content of the problem triggers more sophisticated reasoning strategies. It is possible that the instructional manipulation was not strong enough to facilitate decontextualised reasoning. In the syllogistic reasoning literature the beliefs used are very general and the training focuses on logical necessity which is important for deductive arguments. In the experiments reported in this chapter stereotypes are used which are strongly held beliefs. Klaczynski and Gordon (1996) reported that the motivation to defend beliefs remained strong even when their participants were informed that they would be called back for an interview if their responses were inaccurate. The power of the instructional set may not have been strong enough to engage participants' analytic reasoning systems to override the influence of beliefs. However participants rated the evidence as stronger and more persuasive after these instructions, possibly because they are assuming that the information is true and this removes one reason to provide a more negative evaluation of the evidence. This at least suggests that participants are responding to the instructions, albeit in a different way to what was expected.

In the absence of training, belief-inconsistent problems elicited more sophisticated reasoning strategies than both belief-consistent and neutral problems as in Experiment 2. In contrast to the previous experiment, training led to weaker argument ratings on all the problem types. That is training may have alerted participants to the weakness of the arguments. However, participants still rated arguments which were consistent with their beliefs as stronger than the other arguments.

On the experimental evaluation problems, problems that involved evidence that was inconsistent with beliefs yielded weaker and less valid evidence ratings than consistent or neutral evidence. In accord with the previous experiment, after law of large numbers training participants rated evidence on all the problems as more valid. In Experiment 2 participants rated the evidence as more valid on belief-consistent and neutral problems after training and in this experiment all problems were rated as more valid, whether the evidence was consistent, inconsistent or neutral with beliefs. The control group also rated the strength of the experiments as weaker. This is further evidence that individuals are possibly redirecting their attention after training on the law of large numbers to the wrong components of the problem. They are looking at sample sizes and concluding that they are valid because the sample sizes are quite large.

The following section will provide a general discussion of the findings in relation to both experiments reported in this chapter.

4.10 General Discussion

The findings from the two experiments reported in this chapter provide some support for previous findings in the literature, but also provide evidence that contrasts with the predictions made from dual process theories of reasoning. The aim of the research was to investigate the effects of training on both belief-based and analytic processes used in reasoning. Explicit training on the law of large numbers led to the elimination of biased responding, and the utilisation of more sophisticated reasoning strategies on problems that would normally be processed in a relatively cursory way. Secondly the studies replicated and extended on the findings of Fong et al. (1986; 1991) that training on the LLN principle increases use of statistical reasoning on everyday problems involving different content and structure from the examples used in training, even after a delay. Instructions aimed to foster decontextualised reasoning were found to have no effect. Finally, the results replicated and extended on the work of Klaczynski (Klaczynski, 1997; Klaczynski & Fauth, 1997; Klaczynski & Gordon, 1996; Klaczynski et al., 1997;) by illustrating that problems involving manipulations of belief elicit different reasoning strategies dependent on whether the problem consists of evidence which is consistent or inconsistent with beliefs.

The key finding that training on the law of large numbers increases use of the rule and eliminates bias is a surprising one. This is not fully consistent with any dual process account of reasoning. Whilst it would be clamed that the training may attenuate or reduce the bias, all of the accounts associate belief bias with System 1 so explicit instruction is unlikely to eliminate the influence of belief (Stanovich, personal communication, October 2004; Sloman, personal communication, September 2004; Evans, Newstead, Allen, & Pollard, 1994). Analytic responding was increased dramatically on all problem types, whether they were consistent, inconsistent or neutral to a person's beliefs resulting in an absence of bias.

According to Klaczynski's account of belief effects on these problems, heuristic and analytic responding are independent of each other. Hence training on an explicit rule would not impact on System 1's intuitive system. Klaczynski and Gordon (1996) proposed training should increase the sophistication of reasoning responses on all problems; however the difference between responses on belief-consistent and belief-inconsistent arguments should remain the same. Klaczynski, Gordon, and Fauth (1997) argued that analytic responding was related to measures of intelligence and biases were related to thinking styles. In Klaczynski's view, higher ability participants would acquire the law of large numbers rule more rapidly but they would not be able to utilise it on problems designed to elicit belief-based responses. Klaczynski predicted that rule training would not attenuate the influence of belief on everyday reasoning problems consisting of belief-laden arguments (personal communication, November 2004).

However in contrast to Klaczynski, training increases analytic responding overall and it also reduces the impact of beliefs. The findings show that given training in a rule such as the law of large numbers, people are able to utilise the rule effectively, independently of the influences of beliefs, in generating verbal evaluations of the strength and persuasiveness of the argument.

Interestingly, the effects of beliefs are still present in the rating scales. The first rating scale asked for an evaluation of the strength of the conclusion based on the evidence presented, an evaluation that can be objectively made based upon the characteristics of the samples being discussed. After training the influence of beliefs on this scale is as strong as the influence of beliefs in the control group. This is startling, given that there is no influence of belief on the written justification for these responses. It appears that asking for a simple evaluation of an argument (such as the rating scale) is more susceptible to the influence of System 1 processes. Whereas asking people to generate a written evaluation activates the analytic System 2 processes that make available the LLN principles that have been taught.

The second rating scale asks about persuasiveness of the argument and it could be argued that it is quite rational to be less persuaded by a conclusion that is inconsistent with beliefs. One piece of evidence that is incongruent with beliefs that may often be based on many pieces of evidence, should not in a Bayesian sense impact drastically in changing or persuading us to change our view. If you take a Bayesian approach to the evaluation of evidence then the prior probability of a hypothesis is critically important in determining how to evaluate that evidence and how that impacts on your beliefs. In the case of these problems, the prior probability is the prior belief that you have in the claim, e.g. that nurses are caring. You may have a high degree of belief in that claim and according to Bayes theorem discovering one piece of evidence that is inconsistent with your belief will have a minimal impact on terms of changing that belief. This is particularly the case if it is based on much prior evidence. This type of analysis only applies to the persuasiveness scales because according to Bayes it would be irrational to be persuaded based on one piece of evidence only. However the conclusion strength asked participants to rate how strong they thought the conclusion was based on the evidence used, and in this case prior belief should not be relevant in determining judgement as the response should be based on just the evidence presented.

So why are these results so dramatic considering the predictions that were made? This is the first time that law of large numbers training has been tested on everyday reasoning problems involving belief manipulations. The law of large numbers is proposed to be an intuitive inferential rule system; a rule of thumb that all people possess and can use in everyday reasoning. There is evidence to suggest that LLN reasoning is implicit and also dependent on knowledge of the problem content. Nisbett, Krantz, Jepson, and Kunda (1983) gave participants everyday problems in two domains, sports and acting, which the participants had varied prior experience of. They found that participants with experience in a certain domain were more likely to choose a statistical explanation for the problem whereas if they had no experience they chose a deterministic response. This suggested that people do acquire statistical rules in a domain-specific manner. Support for the conjecture that formal rules for reasoning are induced through experience has been provided by the findings that undergraduates and teenagers use statistical rules to solve problems (Kosonen & Winne, 1995). They do so infrequently and with low levels of sophistication but after formal training on the LLN principle they are able to apply them to different problem formats.

Further support for the view that general instruction in abstract principles improves statistical reasoning is provided by a series of studies which observed the effects of undergraduate training on reasoning (Lehman & Nisbett, 1990; Lehman, Lempert, &

Nisbett, 1988). Lehman and Nisbett reported that undergraduate training in the social sciences resulted in up to 70% improvement in statistical-methodological reasoning. The key finding that training in statistical principles increased the use of LLN reasoning in the domain of training and transfers across into other domains has important implications for dual process account of reasoning. People acquire an intuitive version of the LLN rule tacitly and domain-specifically yet after explicit formal training people were able to utilise the rule on different everyday reasoning problems.

What is it about the training that facilitates this domain general reasoning? According to Fong et al. people are able to map the LLN rules they have learnt onto pre-existing abstract intuitive rules that they can then use on problems in different domains to the one they have been taught. The results here are consistent with that explanation and indeed add more leverage to it as the effects of training were still very strong after a one week delay between training and testing. Fong and Nisbett (1991) found some domain specificity of training with a two week delay but participants who received training still utilised LLN reasoning more on problems in the other domain than participants in the control group.

However, the above findings do not explain why LLN training eliminates belief bias. They focus more on the question of how training increases use of the LLN principle overall. It is possible that people are utilising the rule on the belief motivated arguments as they would on any everyday reasoning problem. The process of using the rule may elicit cognitive decontextualisation on these tasks. Participants have been cued to use the rule which triggers System 2's analytic reasoning strategies. They then read through the problem and identify the small sample size as being a problem, regardless of the conclusion and the direction of belief. Hence when asked for a written evaluation of the evidence participants utilise the rule. However, when they are asked to rate the argument's strength and persuasiveness, their System 1 processes automatically cue the belief-influenced response.

A simple instruction to rate an argument's strength or persuasiveness does not engage System 2's analytic thinking processes.

The alternative explanation is that the reasoning utilised after training is purely superficial and participants in these experiments are simply transferring by analogy. One claim against Fong et al.'s original (1986) study was that the domains used were too narrow therefore the training was not transferring to very different types of problems. Fong and Nisbett (1991) addressed the issue by using broader domains and in these studies the problems used for testing were completely different from the ones used in the training. The examples used in the training in the two experiments reported here were taken from Fong et al.'s objective domain problems but the test items were of a completely different structure and content involving manipulations of belief. Fong and Nisbett proposed that the strong domain independent training effects found when tested immediately after training are due to analogical transfer from the examples used. They explained their domain-specificity of training effects after a two-week delay as being due to recall for the abstract principle not memory for details of the training examples.

One argument against the analogy explanation is that both the experiments reported here support the findings of Fong et al. (1986) that there was only minimal overuse of the LLN principle on problems that were not designed to elicit statistical reasoning. In fact, some overuse of the rule is evidence in itself for a rule-based account of reasoning. According to Smith, Langston, and Nisbett (1992) when acquiring a new rule, it may be overextended and used inappropriately. This is based on a finding in psycholinguistics studies of how children master the regular past-tense form of English verbs. The rule is to add 'ed' to the verb to form the past tense but children tend to overextend the rule to irregular forms e.g. 'give-ed'. Fong et al. included 'false alarm' problems in their study which didn't require a statistical response and concluded that after training, participants were able to use the LLN principle selectively and quite sophisticatedly. Participants only sometimes applied it to cases where it was inappropriate.

No participants in either of the two experiments reported here utilised LLN reasoning consistently on the nine experiment evaluation problems although there was some 'overextension' of the rule noted after LLN training. This leads to the conclusion that individuals were not just learning the rule and applying it to every problem, but were distinguishing between the types of problems and using the rule appropriately. It must be taken into consideration though that the reminder to use LLN reasoning on the statistical problems may have cued participants not to utilise the principle on the experiment evaluation problems.

A second finding which conflicts with the analogy argument is that participants in the training condition in both experiments rated experiment evaluation problems as more valid. It appears that the statistical training may have redirected people's attention to the wrong elements on the problem. In Experiment 2 after training participants' validity ratings increased on belief-consistent and neutral problems. They only pay cursory attention to these problems anyway so statistical training provided a cue for them to identify the sample as a potential flaw. Participants must have identified that the samples were large and concluded that the experiment was valid. In Experiment 3, validity ratings increased on all problem types in experiment evaluation reasoning. However, participants still identified the design flaws on belief-inconsistent problems so levels of reasoning sophistication on these problems did not change after training. In other words, participants were still motivated to find more design flaws on contrary to belief evidence even when they had concluded it was a more valid experiment, probably due to the large sample.

Reeves and Weisberg (1993) also argued against the analogy explanation as they proposed that the recall procedure used by Fong & Nisbett to assess subjects' memory for the exemplars may have led to an underestimation of the amount of exemplar-specific knowledge remembered by the participants. They suggested that a recognition task might have shown that memory for the details of the training exemplars was better than they were led to believe. Reeves and Weisberg argued that the pattern of results obtained by Fong and Nisbett did not support their claims. They claimed that participants referred back to details of examples for help in mapping an abstract principle or formula to a new problem. The findings here are more supportive of Fong and Nisbett's claims as the examples used in training were so different from the ones used in testing.

The finding that instruction had no effect in decreasing bias on either the LLN problems or the experimental evaluation problems was not completely surprising. Klaczynski and Gordon (1996) found that motivation instructions to facilitate more accurate responding had no effect in reducing bias on belief-based problems. It appears that people cannot dissociate from their beliefs consciously and they need more than just motivational instruction. The evidence so far indicates that they need information on the logical rule they must follow. Evans, Newstead, Allen, and Pollard (1994) found a reduction in belief-based responding after an instruction procedure which emphasised the structure of the syllogism and explained the logical meaning of 'some'. In Experiment 3, there was no explanation of the law of large numbers, the rule to utilise on the LLN problems, or any information about experimental confounds built into the instruction set. According to Sloman (personal communication, September 2004) people need to be cued to use a rule, so bias will be reduced only to the extent that the cue is effective. In the instruction condition, there were no cues for either type of problem. This provides further evidence that reasoning after LLN training is purely superficial. Participants are provided with the rule which they utilise on any other problem where it is applicable.

However, more law of large numbers reasoning and more sophisticated experiment evaluation are utilised on belief-inconsistent problems under the instruction and the control conditions. There was no cue present but the content of the problem triggered more sophisticated reasoning strategies. It is possible that the instructional manipulation was not strong enough to facilitate decontextualised reasoning. However there is evidence from Experiment 3 that the instructions are engaging System 2 in some way. Participants rate the evidence strength and persuasiveness as higher which leads us to the conclusion that the instructions are providing them with a reason not to evaluate the evidence strength as low. For instance the instruction to assume the truth of the evidence may be giving participants a reason not to question the truth or accuracy of the evidence. Klaczynski and Gordon reported that the motivation to defend beliefs remains strong even when their participants were informed that they would be called back for an interview if their responses were inaccurate.

What does seem apparent from the previous research and from these two experiments is that we do have some degree of conscious control over our reasoning processes i.e. there are effects of training. The moment-to-moment shifts in reasoning strategies dependent on the content of the problem reported in these two experiments indicated the different reasoning processes being utilised. Klaczynski, Gordon, and Fauth (1997) are proponents of the depth of processing explanation and they suggested that goal-enhancing information was readily assimilated to preexisting belief systems and was processed at a relatively cursory level. On the other hand, goal-threatening evidence activated more sophisticated reasoning tactics and was processed at a deeper level. Klaczynski et al. proposed that these strategies serve to maintain or preserve pre-existing beliefs. The results from the two experiments reported here support this view.

The finding that certain problems elicit use of sophisticated reasoning strategies is not a new one (Nisbett et al., 1983; Jepson et al., 1983). Kahneman and Tversky (1972) argued that people don't take account of sample size on intuitive statistical judgement tasks, however the results here and in previous studies consistently illustrate that people can and do utilise statistical reasoning when certain elements in the problem cue the appropriate response, or when they are trained on the LLN principle. Kahneman and Tversky (1982) revised their argument by proposing that people may understand statistical rules but frequently fail to apply them on everyday problems. To a certain extent this appears to be true. When given everyday reasoning problems in the subjective or objective domain, individuals are very poor at responding using statistical reasoning but when given problems in the probabilistic domain, statistical responses are high. People require the explicit cues about randomness and variability in the problem to guide their response (Jepson et al., 1983).

4.11 Conclusions

By integrating the two bodies of research, the law of large numbers training and the individual differences in everyday reasoning studies, it has been possible to observe the interaction of analytic and belief-based processes. Manipulations of belief within everyday reasoning problems illustrate how individuals' strategies change dependent on whether System 1 or System 2 processes are engaged. The training effects reflect the interactive relationship between the two systems. Under this account, explicit instruction served to trigger the rational System 2 processes which override the implicit System 1 processes leading to the elimination of biased responding and an increase in analytical reasoning.

It is evident from the findings reported that reasoning after training is not superficial. Participants in Experiment 2 utilised the rule provided in the training one week later. They were also able to utilise the rule selectively with only minimal overuse of the rule on different problem types. However, when asked to rate the arguments they indicated their response based on their true beliefs. This demonstrates a dissociation between analytic and heuristic responding on these tasks. When asked to provide a written justification of their reasoning, System 2 processes were engaged and analytic responses given that were not influenced by beliefs. In contrast, when asked to simply rate the argument strength or persuasiveness, participants' responses were cued by System 1 belief-based processes, and showed no influence of training on responses.

The fact that instructions aimed to foster meta cognitive strategies of decontextualisation did not affect reasoning performance at all demonstrates the pervasiveness of beliefs. Evidence in the literature to date illustrates that people find it extremely difficult to separate themselves from their prior knowledge. Giving people the tool with which to reason appears to be a much more effective technique for improving reasoning performance on everyday reasoning problems.

It is difficult to provide an explanation of the findings here based on the predictions made by dual process theories of reasoning as the patterns of results were complex and inconsistent with any of the specific predictions derived from them. One small comfort for these accounts is the dissociation between LLN reasoning and absence of belief effects found in written justifications and the presence of belief effects in the ratings.

Supporters of dual process theories argue about the role general intelligence and thinking styles have in relation to System 2 processes. Stanovich claims that general intelligence and thinking dispositions are closely related to System 2 functioning (Stanovich & West, 2000; Stanovich, 1999). However Klaczynski argues that analytical reasoning is a function of general ability but thinking styles predict bias cued by System 1. Future research should include cognitive ability and thinking dispositions measures to investigate the effects of training or instruction on the responses cued by the belief-based and analytic systems when solving everyday reasoning problems.

In the following chapter a series of experiments will be reported which examine the effects of training on a different type of reasoning problem, the selection task. There is much evidence to suggest that both heuristic and analytic factors influence responding on this task also. The aim is investigate whether the effects of training found in this experiment, specifically in the context of dual process accounts, can be extended to other problem types.

CHAPTER 5

The Effects of Training on the Selection Task

5.1 General Introduction

Chapter 5 presents three experiments designed to investigate the effects of training on the selection task. In Chapter 4 the findings of the two experiments presented illustrated how training on the concept of the law of large numbers increased the quality of reasoning across a variety of problem types. We examined the question of how training moderates the effect of beliefs on behaviour. Experiments 2 and 3 illustrated how people's logical performance can be increased independently from these effects of belief. After training in statistical principles, analytic responding was increased dramatically on all problem types, whether they were consistent, inconsistent or neutral to a person's beliefs resulting in an absence of bias. However, the influence of belief was still present illustrated by the responses on the evidence evaluation and persuasiveness rating scales. Arguments were still viewed as weaker or less convincing if they were contrary to a person's beliefs. In dual process terms, explicit instruction served to trigger the rational System 2 processes which override the implicit System 1 processes leading to the elimination of biased responding and an increase in analytical reasoning, but only when participants were required to justify their response in a written form.

The tasks used in the experiments presented in this chapter are very different from the everyday critical reasoning problems used in Chapter 4. Like the critical reasoning problems, the selection task is a useful tool for investigating pragmatic and analytic influences on correct and incorrect responding. As discussed in the literature review in Chapter 1, there is a great deal of evidence to suggest that correct responding on the deontic selection task is cued pragmatically. However the correct response on the arbitrary or indicative task is obtained with analytic reasoning only. Under a dual process account, System 1 cues correct responding on the deontic task but these cues lead to incorrect responding on the indicative task. System 2 processes on the other hand also lead to correct responding on the deontic task, but if these analytic processes are utilised on the arbitrary task then they are more likely to lead to the correct answer also.

5.1.1 Introduction to Experiment 4

The aim of the experiments presented in this chapter is to see whether training techniques have any effect on responding on the different types of selection task. In Chapter 2 the training studies conducted to date were reviewed. The rest of the introduction in this chapter will concentrate on a more in depth analysis of the training performed and findings reported in relation to the studies presented here. The introduction will end with a detailed rationale for experiment 4.

Experiment 4 was designed to partly replicate and expand on a series of experiments conducted by Cheng, Holyoak, Nisbett, and Oliver (1986) looking at the effects of training on deductive reasoning. As discussed in Chapter 2, they tested two opposing views on how people reason by giving participants training and observing the transfer effects. In one experiment Cheng et al. gave participants logic training in the material conditional and as with Fong's statistical training study, the training was manipulated in that participants were given either training on abstract principles of standard logic, examples of selection task problems followed by an explanation of the correct response or both combined (see Chapter 2 for details of training). They found that training in standard logic, when coupled with training on examples of selection task problems, led to improved performance on subsequent selection task problems. Cheng et al. proposed that abstract principles coupled with examples serve to elucidate the mapping between abstract principles and concrete instances. They found that the rule training alone was ineffective and they proposed that this was because individuals have no ability to apply it to concrete problems. They also found examples training alone to be ineffective and suggest that individuals have no intuitive grasp of the rule they are being shown how to apply. This is very different from Fong et al. who found that rules training alone and examples training alone were effective in increasing use of LLN, and the effects of both together were additive.

Cheng et al.'s rules training consisted of a seven-page booklet containing an in-depth explanation of the logic of conditional statements followed by an inference exercise. Following this participants were given two example problems which involved thematic content. One was the 'Sears' problem; "If the check is $30 or over, then it has to be approved by the section manager" and the second one was "If the painting is cubist, then it is a Picasso". Both conditional statements were embedded in a scenario. Participants were then presented with an explanation of the correct answer for each one written in terms of the abstract rules that they had just learned. It could be argued that Cheng et al.'s logic training was not purely abstract at all. By using examples problems such as this it would be difficult to propose that abstract training in the material conditional transfers to problems in other domains. The training materials could have facilitated transfer because the examples used were thematic in nature.

Eight selection task problems were used to test the training effects: two consisting of arbitrary content, two converse bias, two permission and two arbitrary biconditional problems (see Table 5.1 for the tasks and response choices). The two arbitrary problems were not related to prior knowledge but the converse-bias problems were more realistic and participants' prior knowledge was expected to encourage assumption of the converse (if q then p). The permission problems were expected to be interpretable as a permission schema and therefore elicit responses consistent with the logic of the conditional. Finally the biconditional tasks also consisted of arbitrary content but the problems stated explicitly that the converse of a conditional rule was also true. It was expected that the arbitrary content of the biconditional problems would block application of any pragmatic reasoning schema and that performance on these problems would be poor.

Problem Type	Rule	Choices (*p, not-p, q, not-q*)
Arbitrary	If a card has an 'A' on one side, then it has a '4' on the other side	A B 4 7
	If a bird on this island has a purple spot underneath each wing, then it builds nests on the ground	Bird A has a purple spot underneath each wing Bird B does not have any purple spots Bird C builds nests on the ground Bird D builds nests in trees
Converse bias	If a washing label has 'silk' on one side, then it has 'dry clean only' on the other side	Silk Cotton Dry clean only Machine wash in warm water
	If two objects carry like electrical charges, then they will repel each other	Two objects that carry like electrical charges Two objects that carry opposite charges Two repelling objects Two objects that do not repel
Permission	If a passenger wishes to enter the country, then he or she must have had an inoculation against cholera	Entering Transit Inoculated against cholera and hepatitis Inoculated against typhoid
	If a customer is drinking an alcoholic beverage, then he or she must be over twenty-one	Customer A is drinking a beer Customer B is drinking tea Customer C is certainly over 50 Customer D looks less than 18
Biconditional	If a card has a circle on one side, then it has the word 'red' on the other, and conversely, if it has the word 'red' on one side, then it has a circle on the other	(Picture of a circle) (Picture of a triangle) Red Purple
	If a turtle crosses a road, then the flag by the palace flies, and conversely, if the flag by the palace flies, then a turtle crosses a road	A turtle is crossing a road No turtle is crossing any road The flag by the palace is flying The flag by the palace is not flying

Table 5.1 The rules and corresponding choices in the selection task problems used by Cheng, Holyoak, Nisbett & Oliver (1986)

Cheng et al. found that the training significantly decreased three types of errors on the selection tasks: failure to select p, failure to select not q and erroneous selection of q (see Table 5.2 for the percentage of errors in selection task performance as a function of training however training did not transfer to the solution of biconditional problems. Cheng et al. suggested that if participants had an intuitive appreciation of the material conditional, it might be expected that any advantage gained by training on the conditional would result in some degree of improved understanding of biconditional problems.

	Type of error				
Training condition	*p*	*Not-q*	*q*	*Not-p*	At least one error
Rules & examples	5	27	28	8	39
Rules only	14	48	33	7	65
Examples only	10	45	37	12	62
Control	18	51	44	14	75

Table 5.2 Percentage errors on selection task performance as a function of training condition taken from Cheng, Holyoak, Nisbett & Oliver (1986)

The experimental design and analysis was not sensitive enough to identify the different effects of logic training on arbitrary and deontic selection task problems independently. Only two problems of each type were used which may not be very reliable. Newstead, Handley, Harley, Wright, and Farrelly (2004) found only moderate test-retest reliability correlations on accuracy for the two abstract and two deontic tasks they used (.62 and .34 respectively, overall .38) suggesting random variation in performance on selection tasks. Also Cheng et al. presented no data to show the impact of training directly on each problem type. One of the aims of Experiment 4 was to investigate the effects of training on both arbitrary and deontic tasks independently. In addition we must address the issue of Cheng et al.'s logic training which was not entirely abstract. By ensuring the training is completely abstract and the selection task problems which are used for testing the effects of training are clearly defined, it will be possible to examine the effects of this training on the different types of selection task more clearly.

In another experiment Cheng et al. gave their participants obligation schema training (Cheng, Holyoak, Nisbett, & Oliver, 1986; exp. 3). As discussed in Chapters 1 and 2, it is proposed that people reason using 'pragmatic reasoning schemas', clusters of rules that are acquired through experience (Cheng and Holyoak, 1985). As with LLN reasoning, it should be easy to evoke the use of pragmatic reasoning schemas by presenting individuals with problems consisting of semantic cues designed to trigger them.

In accord with Fong et al.'s suggestion that abstract training in naturally occurring rule systems can be effective in encouraging people to use them, Cheng et al. proposed that if people normally solve problems using pragmatic reasoning schemas then it should be possible to improve people's deductive reasoning by training them on them. The obligation training consisted of details of the nature of obligations in abstract format and the procedures necessary for checking if a violation of the obligation has occurred. An example of an obligation statement presented in the *if-then* conditional form was given. Then the procedures for assessing obligations were described in terms of four rules, one for each of the four possible situations that might arise that can be mapped onto *p*, *not p*, *q* and *not q* (See Chapter 2, Example 2.5 for the full Obligation training script). Two ways that they suggested that schema training could improve performance were firstly by providing participants with more general mapping rules for interpreting situations in terms of the obligation schema and secondly by providing checking procedures consistent with the material conditional that may be applied to other problem types such as those involving arbitrary content.

To establish that it wasn't the checking procedures themselves that resulted in improved reasoning on obligation schema problems, Cheng et al. included a Contingency training condition which involved training participants in the use of checking procedures for contingencies involving the relation between one event or its absence and another event or its absence (see Table 5.3 for Contingency training).

Contingencies

A *contingency* arises whenever it is the case that a certain condition implies some necessary consequence. Contingencies can be stated in an "*If ....then*" form. For example, the following statement specifies a contingency: "If a mushroom is red, then it is edible." Another example would be "If a student is a psychology major, then the student must take an introductory psychology course." More generally, if we call the initial condition *I* and the consequence *C*, a universal contingency has the form, "*If I, then C.*" In our first example, *I* is "red mushroom" and *C* is "edible." In the second example, *I* is "being a psychology major," and *C* is "taking an introductory psychology course."

In order to assess whether a contingency in fact holds, we need to consider the four possible situations that might arise. These are

1. *I* is obtained
2. *I* is not obtained
3. *C* is obtained
4. *C* is not obtained

Corresponding to each of these possible situations is a rule related to the truth of the contingency. These rules are the following:

1. If I is obtained, then C must be obtained. Clearly, if I is obtained then for C not to obtain would show that the contingency doesn't hold. To use our examples, if a mushroom is red, then it must be edible or else the contingency is false: and if a student is a psychology major, then the student must take an introductory psychology course.

2. If I is not obtained, then the contingency is not tested. C need not obtain, although it may. For example, if a mushroom is not red, it need not be edible, although it may be (perhaps brown mushrooms are also edible). Similarly, if a student is not a psychology major the student need not take an introductory psychology course. It may be possible however, for an English major to take an introductory psychology course. But in any case, the basic contingency is simply irrelevant if the student is not a psychology major.

3. If C is obtained, then the contingency is certainly not falsified, regardless of whether or not I obtains. If I didi occur, then the contingency is satisfied. If I is not obtained, then the contingency wasn't even tested (Rule 2). For example, if we know a certain mushroom is edible, we can be sure the contingency was not falsified regardless of the mushroom's colour. Either the mushroom is red, and the contingency is satisfied, or it is not red, in which case the contingency was not tested. Similarly, if we know a student has taken an introductory psychology course, we can be sure the contingency has not been falsified: Either the student was a psychology major, and hence satisfied the contingency, or the student was not a psychology major, in which case the contingency wasn't tested.

4. If C is not obtained, then I must not obtain or else the contingency is falsified. This is because if I had occurred, then the failure to obtain C would falsify the contingency. Then is a mushroom is not edible, it must not be red or else the contingency is false. A nd is a student has not taken an introductory psychology course, the student must not be a psychology major, or else again the contingency will be falsified.

If you understand the above four rules, you should find it easy to assess whether or not a contingency is being satisfied. Note that there are only two situations in which it is possible for a contingency to be falsified: When I is obtained (and C is not obtained) (Rule 1), and when C is not obtained (and I is obtained) (Rule 4). In the other two situations the contingency can't be falsified. These are the cases in which I does not obtain (in which case the contingency will have been satisfied it it was tested) (Rule 3).

You may wish to reread these instructions carefully in order to be sure you understand the rules for evaluating contingencies. You will then be able to apply what you learned to the test problems, which will include a variety of such contingencies. You will find it easy to solve these problems if you carefull apply Rules 1-4.

Table 5.3 Contingency Training taken from Cheng, Holyoak, Nisbett, & Oliver (1986; Exp.3)

They expected that teaching participants correct checking procedures in an abstract form would aid the solution of arbitrary problems only and have little effect on obligation problems. This was due to the fact that these semantically interpretable problems would be understood in terms of a reasoning schema that maps onto the conditional therefore checking procedures would be redundant. If not these problems would be understood in terms of a reasoning schema that does not map onto the conditional therefore would have its own checking procedures that would override the arbitrary procedures.

Participants were given four arbitrary problems and four obligation problems for the test. Cheng et al. reported that obligation schema training improved performance on obligation problems more than contingency training or a control group who received no training. Also as Cheng et al. expected, contingency training resulted in no improvement on problems consisting of obligation content but it did facilitate some improvement on the arbitrary problems. Cheng et al. also found that the benefit of obligation training extended to two of the arbitrary problems and when they investigated this further they found that response patterns under each condition on two of the problems resembled those of obligation problems. They concluded that some participants might have interpreted the two arbitrary problems as obligation situations after training.

Cheng et al. reported which two arbitrary test problems may have been interpreted as obligation schema problems however on examination of them it is not surprising. The two that it is suggested were possibly interpreted as obligations were; 'if a bolt of cloth has any red threads in it, then it must be stamped with a triangle' and 'if a house was built before 1979, then it has a fireplace'. It could be argued that both these tasks are highly thematic and much more likely to be understood as deontic conditionals rather than ones that are purely arbitrary in content. Therefore it is not clear from Cheng et al.'s findings whether training in an obligation schema does transfer to arbitrary problems.

Klaczynski, Gelfand, and Reese (1989; Klaczynski & Laipple, 1993) would probably propose that transfer did not occur from obligation schema training to arbitrary tasks. Klaczynski, Gelfand, & Reese (1989) conducted a series of experiments to observe the transfer effects between thematic and arbitrary selection task problems. They found that transfer occurred within problem types i.e. arbitrary to arbitrary, thematic to thematic, but transfer between problem types was unidirectional. That is transfer occurred from arbitrary to thematic problems only. Klaczynski et al. (1989; Klaczynski & Laipple, 1993) proposed that when given training in an abstract domain, the rules are relatively general because there is no concrete knowledge base onto which the rules can be mapped. Therefore, once the rule is induced, transfer may occur to different problem types that are perceived as instantiations of the general category of problems to which the rule applies. According to Klaczynski and Laipple, because Cheng et al.'s rules training is likely to be consistent with the rules of the deontic problems, the training may be directly mapped onto and activate the schemas which may then be applied to other problems.

However, Klaczynski et al. found no transfer from permission schema problems to arbitrary tasks. They proposed that training in a domain-specific rule may lead to the rule being utilised on other problems from the same domain as one-to-one mapping of the elements between problems is more readily perceived, but transfer to problems in other domains will not occur as the rule is tied to a specific domain. The training should result in activation of that schema which will facilitate the solution of further problems within that domain with little or no transfer to problems from other domains.

Ansberg and Shields (2003) provided support for Klaczynski et al.'s findings by showing that feedback in the form of domain-specific explanation did not promote transfer from permission problems to arbitrary problems. However when participants were required to perform problem comparisons and read strategy instructions during practice the solution procedures did transfer to arbitrary problems. They proposed that this kind of training emphasised the deep structure of the problems by teaching solvers to identify the important problem elements, regardless of the context. This is in contrast with the problem-specific instruction procedure which emphasised the surface structure of the permission problems only.

In terms of dual process theories it is System 2 that is sensitive to instruction. System 2 is associated with intelligence and analytic reasoning therefore under this view explicit training could transfer to other types of problems where the same elements of the problem may be mapped on and lead to the same conclusion, for instance from deontic obligation schemas to arbitrary tasks, but only for individuals who have the ability to decontextualise the key elements from the context. Findings in relation to cognitive ability and thinking dispositions, and claims made by Klaczynski and Laipple (1993) may help in our understanding of the conflicting findings reported so far.

It is proposed that intelligence is fixed and thinking styles are malleable and it is the more open-minded flexible thinkers that are more sensitive to training (Stanovich, 1999). According to Stanovich (Stanovich, 1999; Stanovich & West, 1998b), individuals of a higher ability are able to flexibly contextualise and decontextualise information dependent on the demands of the task. To transfer between problem types individuals must be able to separate elements in the problems from the scenarios they are embedded in and map them onto other problems. This may be easier from abstract training or arbitrary problems as the elements are not so embedded in the first place.

Klaczynski and Laipple (1993) investigated the role of intelligence when transferring domain-independent, arbitrary and domain-specific, deontic knowledge to target domain-independent and specific problems. They found only weak relationships with ability when transferring from domain-independent problems as the rule may be more easily applied to target problems. They reported strong relationships with ability when transferring knowledge from domain-specific problems and they suggested that individuals with higher ability are able to perceive the underlying structural similarities between different problems and therefore apply the same rules to each. It must be noted that Klaczynski reported no transfer from permission schema training to problems from other domains.

Stanovich (1999) also proposes that it is high ability individuals that are able to inhibit System 1 (p/q) responding on the arbitrary selection task. It is these individuals that are able to override the default responses from System 1, the fundamental computational bias. Evidence in support of this has been provided by Houdé and Moutier (1996; 1999) who illustrated that inhibition training successfully led to inhibition of the heuristic response. In addition the area of the brain being utilised after training was in fact related to working memory, which is an indicator of System 2 processing.

Stanovich (1999), Klaczynski, Gordon, and Fauth (1997) and Houdé et al. (1999) all provide evidence to show that thinking styles are also an important part of System 2 analytical thinking. Stanovich proposes that they are linked to cognitive ability and both together predict decontextualised reasoning. In contrast Klaczynski et al. argue that decontextualised reasoning is completely separate from intelligence. Intelligence predicts quality of reasoning but bias is predicted by thinking styles. By including individual difference measures in this experiment, it will be possible to investigate their mediating effects on selection task reasoning and to observe how they affect sensitivity to training.

5.1.2 Rationale for Experiment 4

Experiment 4 was designed to compare Cheng et al.'s rules plus examples training and obligation schema training. Cheng et al. looked at these training techniques in two separate experiments. By investigating both in the same study, direct comparisons between both types of training can be made. The main aim of the experiment is to observe the effects of explicit training in 'domain-general' rules and 'domain-dependent' schemas on subsequent selection task problems involving arbitrary and deontic content. Clearly defined selection tasks that consist of arbitrary and deontic content will be employed to test the differential training effects.

The findings in relation to the previous training studies are not clear. According to Cheng et al.'s findings, it is proposed that abstract logic training serves to improve overall performance on selection tasks. Firstly their abstract training was not entirely abstract. Both of the examples that were used involved thematic content and could be interpreted as deontic problems. In order to make claims about cross domain transfer from an abstract training procedure, then all the training must involve abstract content. The examples used in Experiment 4's training were therefore changed to problems consisting of arbitrary content.

The obligation training was also slightly adapted to produce a procedure that was more equal in length and structure to the rule-based training. Namely, two obligation selection task problems were given after the details on obligation training, each followed by full explanations of the correct and incorrect responses written in the same terms as the initial training. Cheng et al. reported that the obligation training may have refined participants' understanding of situations that are potentially interpretable in terms of the schema. They based this claim on the evidence that obligation schema training transferred to two arbitrary problems. However the two problems in question were highly thematic and could have been interpreted as obligations quite easily. The arbitrary test problems in Experiment 4 have been used in previous experiments and have been shown to elicit patterns of response consistent with abstract selection task problems. By replicating the training and including more reliable test problems, it is possible to more reliably test the transfer effects, at the same time enabling the clarification of the conflicting arguments between Cheng et al. and Klaczynski et al. To recap, according to Cheng et al. transfer does occur from domain-specific schema training but according to Klaczynski. et al. transfer does not occur.

Another aim of this experiment is to investigate the mediating effects of individual differences in cognitive ability and thinking styles. Cheng et al. did not include individual differences measures in their studies therefore by using them in this experiment it may again help explain some of the conflicting findings they had with Klaczynski et al. Klaczynski and Laipple proposed that the ability to transfer from schema-based training is related to ability therefore it is possible that the transfer effects found from Cheng et al.'s obligation training were due to very high ability participants. By including cognitive ability and thinking styles measures it will be possible to explore the relationships that these factors have with the ability to reason successfully on these tasks before training, and also the ability to transfer knowledge after training.

Under dual process accounts there are two different predictions to be made in relation to performance after the abstract logic training. Evans (1989) argued that few people get past a relevance judgement on the selection task and suggests that it is doubtful that verbal instruction in relation to underlying logical principles can remove biases as they are implicit. In other words, explicit verbal instruction would have an impact on verbal and explicit thought processes, not implicit ones. However Klaczynski et al. (1989; Klaczynski & Laipple, 1993) would predict that a domain-independent rule would be induced from the training that would then be applied to both the arbitrary and deontic problems.

Evans may make the same prediction about the effects of obligation schema training as for the logic training. On the other hand, if the training can refine people's understanding of situations in terms of the schema as Cheng et al. proposed, then there will be improved reasoning performance on permission problems and arbitrary problems also. However, according to Klaczynski et al. the training should improve reasoning on problems designed to elicit the obligation schema only.

In terms of the individual differences measures, Sloman (personal communication, September 2004) would predict that only the high ability participants would be able to transfer from both types of training as it is those individuals that are able to understand the training. Stanovich would predict that correlations between individual differences in ability and performance would attenuate after training (personal communication, October 2004) and Klaczynski et al. (1993) would agree with this but only for the logic training. In their view, because obligation schema training is domain-specific, only individuals of higher ability will recognise the structural similarities between the different problems resulting in positive transfer to other domain-specific problems.

Stanovich (1999), Houdé et al. (1999) and Klaczynski et al. (1997) would agree that thinking dispositions are predictive of people's receptiveness to training. It would be expected that these participants would be more "rational" in their style of thinking as measured by the Rational-Experiential Inventory (Pacini & Epstein, 1999) therefore be more able to engage in sophisticated thinking. It would also be expected that these participants would score more highly on Stanovich's Thinking Disposition Questionnaire (see Chapter 3) indicating the availability of a larger repertoire of reasoning styles that they can use to flexibly reason with depending on the context of the problem.

.

5.2 Method

Design

The experiment had a between subjects pre-test/post-test design consisting of two conditions based on Cheng, Holyoak, Nisbett, and Oliver (1986), Rules plus examples training (adapted from Cheng et al. (1986:Experiment 1)) and Obligation Schema training (adapted from Experiment 3). The dependent variables were eight selection task problems (four arbitrary, four deontic) and eight conditional reasoning problems. The two training conditions had not been directly compared before.

Participants

Sixty participants, 35 female and 25 male undergraduate and postgraduate students from the University of Plymouth were randomly allocated to one of the two training conditions (obligation group, 14 females and 16 males; rules group, 21 females and 9 males.). The mean age of the sample was 24.2 years (obligation group 24.03, rules group 24.37).

Materials

Rules plus Examples Training

This training condition was adapted from Cheng et al.'s Experiment 1 (See Appendix A4.1 for complete script). In their experiment they had three training conditions (plus control), Rule Training, which constituted a booklet consisting of an explanation about conditional statements, followed by an inference exercise. Examples training which consisted of two thematic selection task problems with immediate feedback on performance, and Rules plus Examples training, which were the two previous conditions added together.

The Rule training consisted of an explanation of the equivalence between a conditional statement and its contrapositive, as well as an explanation of the two common fallacies of affirming the consequent and denying the antecedent. The contrapositive was explained in part by the use of a truth table, in part by Euler diagrams that used concentric circles to show the relations between a conditional statement and its contrapositive, and in part by an illustrative conditional statement. For this experiment, the section on Euler diagrams was omitted.

Participants were then given an inference exercise consisting of three conditional statements involving realistic content. They had to select statements that could be validly inferred from each of three given conditional statements. The statements were all in the form of *if p then q*. For example, if the cube is plastic, then the sphere is metallic. The randomly ordered possible inferences were in the following forms: *if not-p then not-q* (invalid) if the cube is not plastic, then the sphere is not metallic, *if not-q then not-p* (valid) if the sphere is not metallic, then the cube is not plastic, and *if q then p* (invalid) if the sphere is metallic, then the cube is plastic. Participants made their selection and the correct answer was given on the following page.

Participants were then presented with two selection task problems. In Cheng et al.'s study the examples involved thematic content. For this study they were changed to arbitrary selection tasks in order that the condition was completely abstract. Following each problem, participants were given a full explanation of the correct and incorrect card choices in terms of *p, q, not p* and *not q*.

Obligation Schema plus Examples Training

This training condition was adapted from Cheng et al.'s Experiment 3 (See Appendix A4.2 for complete script). Their obligation training consisted of a two-page booklet detailing the nature of obligations and the procedures necessary for checking if a violation of the obligation had occurred. An example of an obligation statement presented in the *if-then* conditional form was given. The procedures for assessing obligations were described in terms of four rules, one for each of the four possible situations that might arise – situations that can be mapped onto *p*, *not-p*, *q*, and *not-q*. In other words situations where p occurs, p does not occur, q is done, and q is not done.

For this experiment two examples were also included which were designed to elicit the obligation schema. Participants were asked to complete each problem before checking the answer on the following page. A full explanation detailed the correct and incorrect card choices that could be made to check if there has been a violation of the obligation rule. Examples weren't used in the original study but in order that direct comparisons with the other training condition could be made, they were included here.

Selection Task Problems

Sixteen selection task problems were used in total, 8 abstract, 4 obligation and 4 permission. Eight problems were used as the pre-test and eight after training as the post-test. The problems that constituted the pre-test for half of the participants, became the post-test problems for the other half, and the post-test problems became the pre-test. There was no difference between forms 1 and 2 on the arbitrary tasks prior to training ($t(58) = 0.00$, $p=1$) but there was a significant difference between the two forms on the deontic tasks at $t(58) = -2.06$, $p<.05$. All were presented in random order.

The eight abstract problems consisted of the original letter and number task (Wason, 1966), the destination problem used in Experiment 1 and others adapted from Klaczynski and Laipple (1993). All were designed in order that responses would not be affected by prior knowledge or belief. Participants were given the rule then asked which card or cards they would need to turn over to see whether the rule is true or false. Response orders were varied between problems (see Table 5.4 for the conditional statements and response choices).

Form	Conditional statement	Response
Form 1	"If the letter A is on one side of a card, then the number 7 must be on the other side of the card."	A (p)
		D (not p)
		7 (q)
		3 (not q)
	"If Triangle is on one side of a card, then Dog must be on the other side of the card."	Triangle (p)
		square (not p)
		dog (q)
		cat (not q)
	"If Yellow is on one side of a card, then Oak must be on the other side."	yellow (p)
		green (not p)
		oak (q)
		beech (not q)
	"If Coffee is on one side of a card, then Goldfish must be on the other side of the card."	coffee (p)
		tea (not p)
		goldfish (q)
		hamster (not q)
Form 2	"If Glasgow is on one side of a card, then Train must be on the other side of the card."	Glasgow (p)
		Edinburgh (not p)
		train (q)
		bus (not q)
	"If Salty is on one side of a card, then Pistachio nut must be on the other side of the card."	salty (p)
		sour (not p)
		pistachio nut (q)
		walnut (not q)
	"If Table is on one side of a card, then Daisy must be on the other side of the card."	table (p)
		chair (not p)
		daisy (q)
		buttercup (not q)
	"If Trousers is on one side of a card, then French must be on the other side of the card."	trousers (p)
		skirt (not p)
		French (q)
		Italian (not q)

Table 5.4 Conditional Statements and Response Choices for the Arbitrary Selection Task Problems

The four obligation selection task problems described situations in which the occurrence of some condition A incurs the necessity of taking some action B. Three problems were adapted from Cheng et al. (Experiment 3) and one was created by the author. The four permission selection task problems contained rules describing an action that could occur if a precondition had been met. Three of these had been created by the author and one was the drinking-age problem from Experiment 1. For both types of problems participants were asked to indicate which card or cards they would need to turn over to decide whether or not the rule was being violated (see Table 5.5 for the conditional statements and response options). Again, response orders were varied between problems.

Problem type and Form	Conditional statement	Response
Obligation - Form 1	"If a urithium miner gets lung cancer, then the company must pay the miner a sickness pension."	Miner B has lung cancer (p) Miner D does not have lung cancer (not p) Miner A is receiving a sickness pension (q) Miner C is not receiving a sickness pension (not q)
	"If a steel support is intended for the roof, then it must be rustproof."	Support A is intended for the roof (p) Support B is intended for the foundation (not p) Support C is rustproof (q) Support D is not rustproof (not q)
Obligation - Form 2	"If a person is driving a car, then that person must wear a seatbelt."	Driving a car (p) Not driving a car (not p) Seatbelt worn (q) Seatbelt not worn (not q)
	"If one works for the Armed Forces, then one must vote in the elections."	Person B does work for the Armed Forces (p) Person A does not work for the Armed Forces (not p) Person C voted (q) Person D did not vote (not q)
Permission - Form 1	"If a person is travelling to the United States of America, then that person must have a visa."	United States of America (p) France (not p) Passport and a Visa (q) Passport and no Visa (not q)
	"If a person is drinking beer, then that person must be at least 18 years old."	Drinking beer (p) Drinking Coke (not p) 18 years of age (q) 16 years of age (not q)
Permission - Form 2	"If a student is withdrawing books from the library, then they must have a valid Identity card"	Books on loan (p) Books not on loan (not p) Valid Identity card (q) Invalid Identity card (not q)
	"If a person purchases cigarettes or tobacco, then they must be 16 years or over in age."	Bought cigarettes (p) Didn't buy cigarettes (not p) 17 years of age (q) 15 years of age (not q)

Table 5.5 Conditional Statements and Response Options for the Deontic Selection Task Problems

Conditional Reasoning Problems

Sixteen conditional reasoning problems were used in total, eight pre-test and eight post-test. Prior to each test booklet, participants were given a set of instructions (see Appendix A4.3 for the instructions and the problems) asking them to decide which of the three conclusions that followed the premises were valid, i.e. those which logically have to follow given that the premises are true. All the logical problems involved abstract content, two designed to draw modus ponens inferences, two modus tollens, two denial of the antecedent and two affirmation of the consequent. The three options were q, not q and 'impossible to tell' for the MP and MT inferences; and p, not p and 'impossible to tell' for the AC and DA inferences. All the problems were presented in random sequence.

The items were scored on whether participants had obtained the logically correct response, which involved q for MP, not p for MT, and 'impossible to tell' for the AC and DA inferences. Conditional reasoning problems were included to investigate whether the rules of standard logic training or obligation schema training would transfer to a set of conditional reasoning items.

Cognitive Ability

The AH4-Group test of General Intelligence was administered (Heim, 1967) as used in Experiment 1. The test consists of two parts, each containing 65 items. Part one is comprised of both verbal and numerical problems, part two is composed of problems in diagrammatic form. Correlations between scores on Part One and scores on Part Two have been reported to range between 0.60 to 0.81 (Heim, 1967) and in Experiment 1 the correlation was 0.56. Each section is timed and must be completed in 10 minutes.

Thinking Dispositions

The Rational-Experiential Inventory and Stanovich's Thinking Dispositions Questionnaire were used. See Experiment 1 for details.

Procedure

Participants were tested in groups of two to six. They were given a booklet which contained all the test materials and training to work through. First they completed the timed component of the test, the AH4, after which they worked through the rest of the booklet in their own time. They completed the pre-test conditional reasoning problems followed by the eight selection tasks. Participants were then requested to read through the training materials and make sure they understood before progressing to the remainder of the booklet. Following the training, participants completed the Rational-Experiential Inventory and Stanovich's Thinking Disposition Questionnaire before completing the target selection task and conditional reasoning problems. The whole task took an hour to complete.

5.3 Results

5.3.1 Selection task responding under each training condition

Responses on the selection task problems were scored as 1 for a *p/not q* (correct) response and 0 for any other (incorrect) answer. Due to the unequal number of tasks under each problem type: two obligation problems, two permission problems and four arbitrary problems, the mean proportion of correct answers was computed for each problem type. Table 5.6 illustrates the mean pre and post training scores on the two tasks designed to evoke an obligation schema, the two tasks designed to evoke a permission schema and the four tasks consisting of arbitrary content. On initial observation the pre training scores on the obligation and permission problems do not differ much but a t-test confirmed there was a significant difference between the two at $t(59) = -2.82$, $p<.01$ with permission schema problems yielding a higher number of correct answers than obligation schema problems.

	Obligation training		Rules training	
Problems	Pre-test	Post-test	Pre-test	Post-test
Obligation	.48	.75	.50	.52
Permission	.65	.80	.60	.58
Arbitrary	.07	.21	.13	.39

Table 5.6 Mean proportion of correct answers on the pre-test and the post-test for the selection tasks (N=30 per training group)

Illustrated in Table 5.6 are the scores on the selection tasks pre and post training. Pre-test scores on both the deontic tasks are higher than pre-test scores on the arbitrary tasks which is supportive of Cheng et al.'s reasoning schema hypothesis and all the other findings in the selection task literature (see Evans, Newstead & Byrne, 1993). The deontic problems were designed to evoke the use of the permission and obligation schemas resulting in more correct responses on these problems. A t-test showed the difference between scores on the obligation and arbitrary problems to be significant ($t(59) = 7.26$, $p<.001$), and the difference between scores on the permission and arbitrary problems to be significant ($t(59) = 9.03$, $p<.001$). A 2 x 3 x 2 (Condition x Problem Type x Pre-Post test scores) ANOVA was performed on the data with Condition as between subjects and Problem type and pre-post test scores as within subjects variables. The ANOVA failed to find a main effect of condition ($F(1, 58) = p>.1$), but there was a main effect of pre to post training scores ($F(1, 58) = 8.85$, $MSE = 1.70$, $p<.01$). No significant interaction between training condition and pre to post training scores was obtained ($F(1, 58) = 1.24$. $MSE = 0.24$, non-significant). There was a main effect of problem type ($F(2, 116) = 68.11$, $MSE = 6.94$, $p<.001$) and an interaction between training condition and problem type ($F(2, 116) = 5.77$, $MSE = 0.59$, $p<.01$). An interaction between problem type and pre to post training scores failed to reach significance ($F(2, 116) = 2.44$, $MSE = 0.14$, non-significant).

The ANOVA yielded a significant interaction between Condition, problem type and pre-post test scores ($F(2, 116) = 4.53$, $MSE = 0.26$, $p<.05$). This interaction is illustrated in Figure 5.1. Participants achieved higher scores for each problem type after obligation training. In the rules training condition higher scores were achieved for the arbitrary problems only ($F(1, 58) = 13.85$, $MSE = 1.00$, $p<.001$). The difference pre to post training on the obligation and permission problems was not significant ($p>.1$ for both). A series of planned comparisons revealed that under the obligation schema training the difference pre to post scores for the obligation problems was significant ($F(1, 58) = 6.65$, $MSE = 1.07$, $p<.05$), for the permission problems the difference was significant ($F(1, 58) = 4.44$, $MSE = 0.34$, $p<.05$), and for the arbitrary problems the difference was significant ($F(1, 58) = 4.67$, $MSE = 0.34$, $p<.05$). In sum, under obligation schema training performance was improved for arbitrary and deontic tasks and under rules training performance was improved for arbitrary problems only. There was no transfer to obligation or permission problems after rules training.

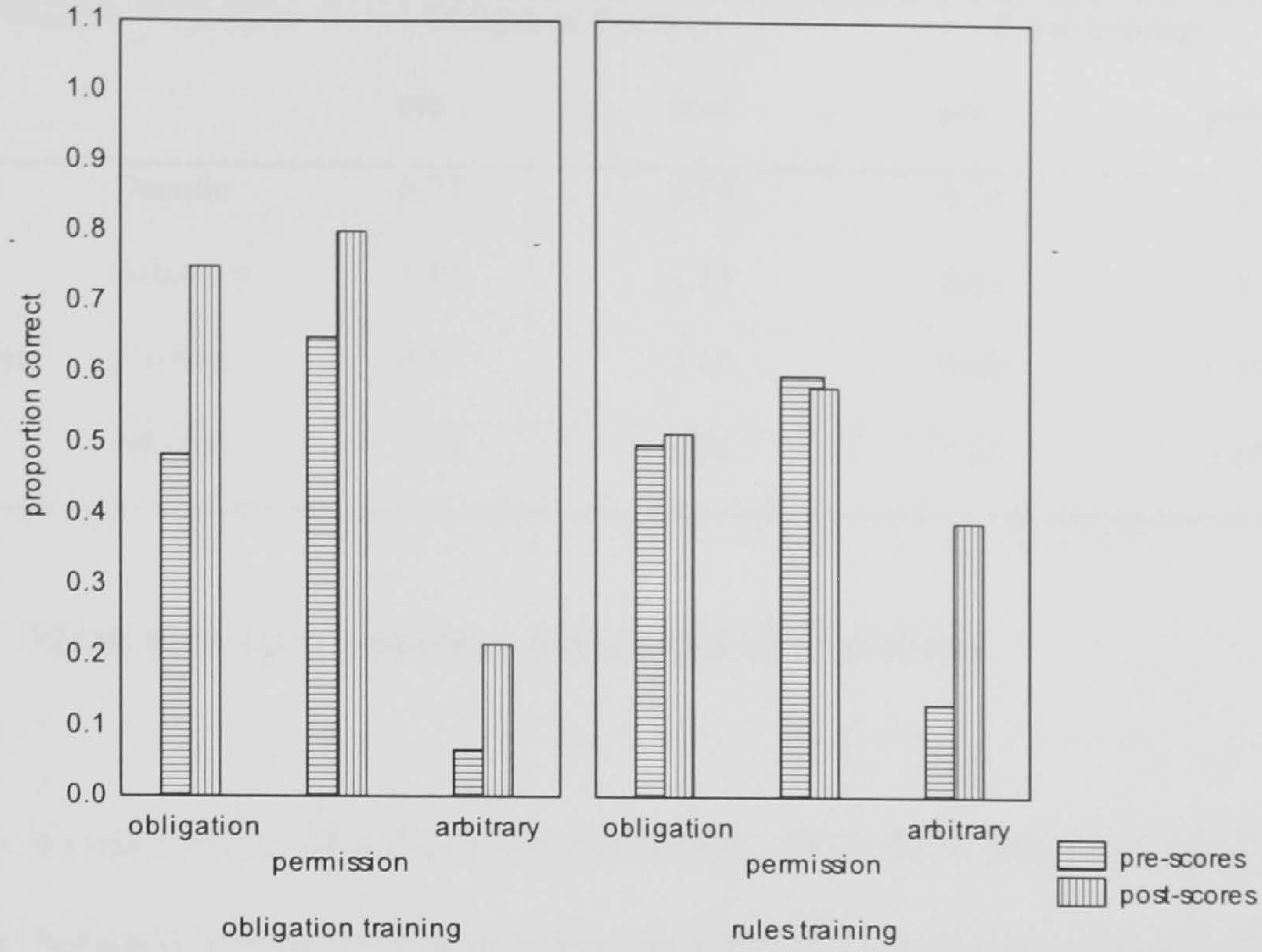

Figure 5.1 Graph illustrating 3-way interaction between Condition, Problem type and pre and post training scores (N = 30 per training group)

5.3.2 Changes in response patterns

The pattern of responding on the obligation and permission problems was the same after both types of training therefore they were collapsed into one problem type, deontic problems, for further analysis. The following analysis was an attempt to investigate the changes in response patterns dependent on the type of training. In other words, an investigation into how the different types of training affected responding directly. As the most common responses on the arbitrary selection tasks in the absence of training are p and p/q, these patterns were the ones explored (see Table 5.7 for means). Participants' performance before and after training on the two problem types, deontic and arbitrary were scored for p responses only. A 2 x 2 x 2 (Condition x Problem Type x Pre/Post scores) ANOVA was performed. The ANOVA failed to find an effect of condition ($F(1, 58) = 3.31$, $MSE = 11.70$, $p>.05$) although it did find a main effect of problem type ($F(1, 58) = 8.80$, $MSE = 14.50$, $p<.01$) with a higher rate of p responding on arbitrary problems than deontic problems overall.

		Obligation training		Rules training	
		pre	post	pre	post
P responses	Deontic	0.77	0.27	0.20	0.57
	Arbitrary	1.50	1.13	0.60	0.53
P/q responses	Deontic	0.17	0.17	0.90	0.70
	arbitrary	1.70	1.60	2.23	1.07

Table 5.7 Mean p and p/q responses before and after training

There was no main effect of pre to post scores (F(1, 58) = 0.90, MSE = 1.20, p>.05). The interaction between condition and problem type failed to reach significance (F(1, 58) = 3.46, MSE = 5.70, p>.05) however the interaction between condition and pre to post scores was marginally significant (at F(1, 58) = 3.83, MSE = 5.10, p=.05). Figure 5.2 illustrates that in the obligation training condition participants' p responding decreases (F(1, 58) = 4.22, MSE = 5.63, p<.05) however under rules training p responding increases, although not significantly (F(1, 58) = 0.50, MSE = 0.67, p>.1).

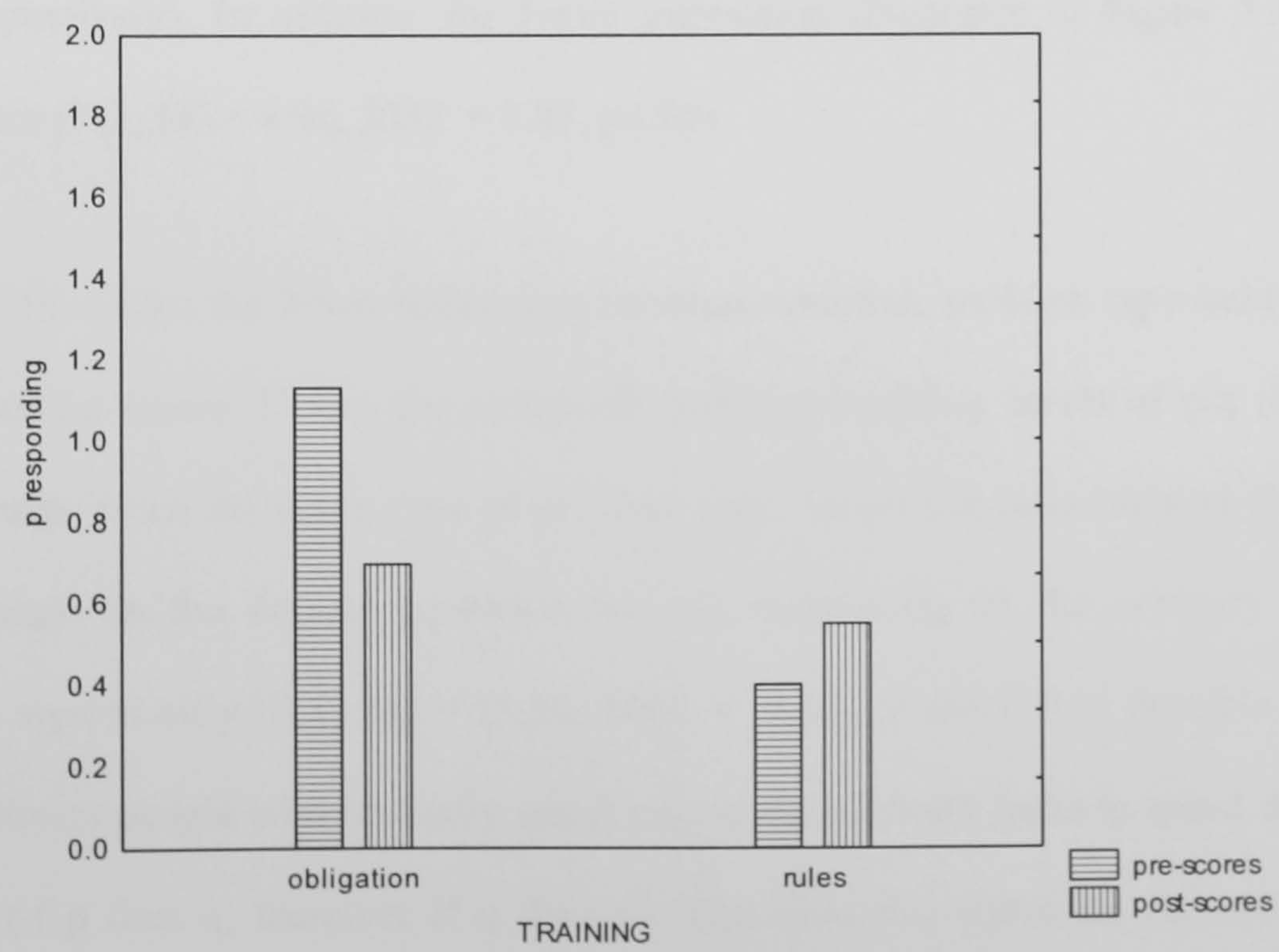

Figure 5.2 Interaction between condition and pre to post scores on p responding

The interaction between problem type and pre to post scores also failed to reach significance (F(1, 58) = 0.72, MSE = 0.34, p>.1) as did the 3-way interaction between condition, problem type and pre to post scores (F(1,58) = 2.57, MSE = 1.20, p>.1).

Participants' performance on the two types of selection task problems were then scored for p/q responses before and after training. Another 2 x 2 x 2 (condition x problem type x pre/post scores) ANOVA was performed to investigate changes in responding dependent on training type. No main effect of condition was obtained (F(1, 58) = 1.32, MSE = 6.02, p>.1) but there was a main effect of problem type (F(1, 58) = 38.41, MSE = 81.67, p<.0001) with higher p/q response rates on arbitrary problems than deontic. There was also a significant main effect of pre to post training (F(1, 58) = 7.81, MSE = 8.07, p<.01) with p/q responses being decreased after training. The condition by problem type interaction failed to attain significance (F(1, 58) = 2.83, MSE = 6.02, p>.05) however both the condition by pre to post training and the problem type by pre to post training interaction were significant (F(1, 58) = 5.82, MSE = 6.02, p<.05 and F(1, 58) = 7.52, MSE = 4.27, p<.01 respectively). In addition the 3-way interaction illustrated in Figure 5.3 attained significance (F(1, 58) = 4.96, MSE = 2.82, p<.05).

Figure 5.3 illustrates the 3-way interaction between condition, problem type and pre to post p/q responding scores. Under the obligation training condition levels of p/q responding didn't change much for either type of problem type. Under the rules training there is not much change on the deontic problems but p/q responding on the arbitrary problems decreases significantly (F(1, 58) = 13.56, MSE = 14.01, p<.001). It is possible that rules training directs people who normally select p/q on the arbitrary tasks to avoid the inverse response (if p then q, therefore if q then p). This does not affect the deontic problems because p/q responses were low prior to training anyway.

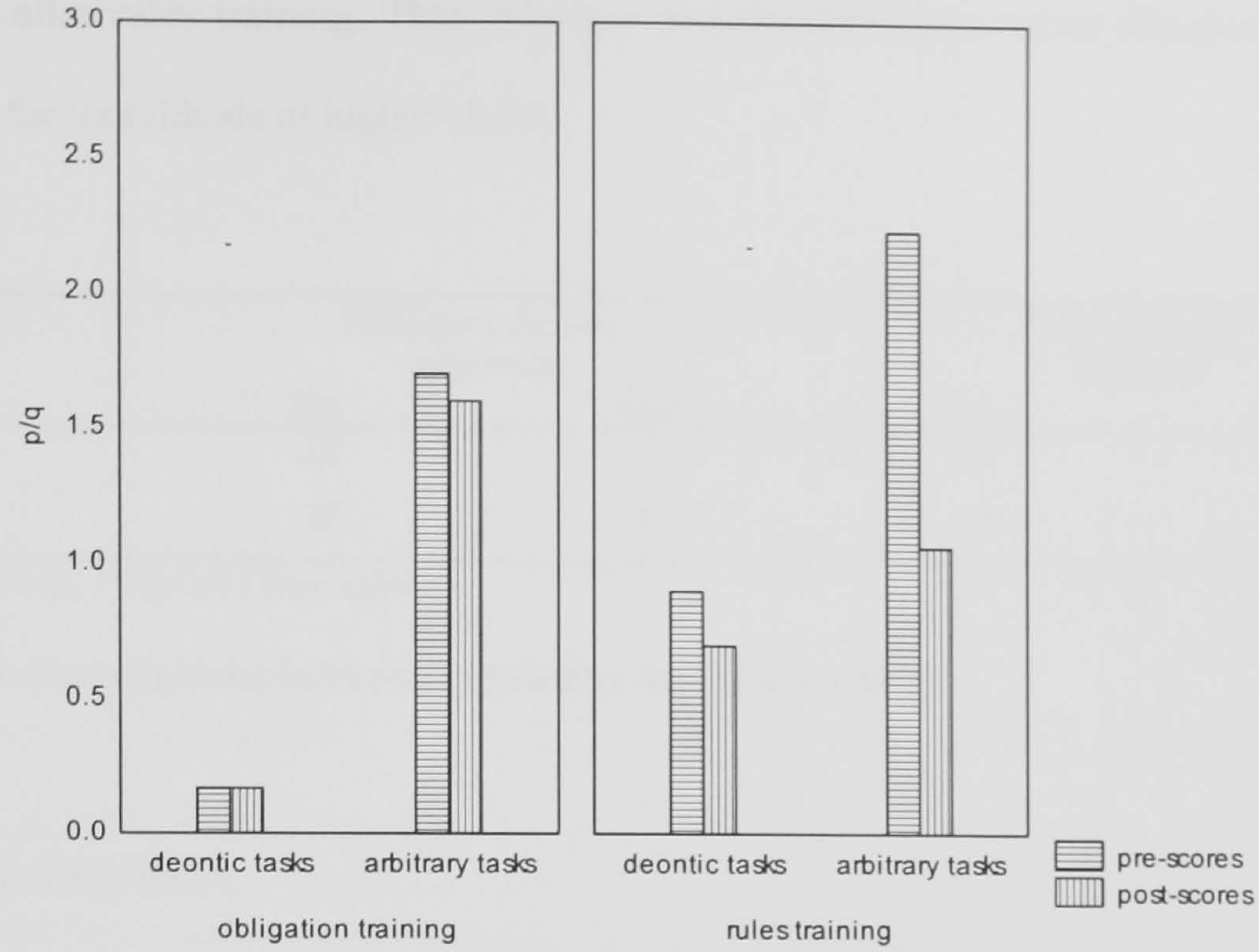

Figure 5.3 3-way interaction between condition, problem type and pre to post p/q responding scores

5.3.3 Cognitive ability

Correlations were performed on the data to observe whether there were any associations between performance on the reasoning tasks and cognitive ability (see Table 5.8). The mean proportion of correct items per problem type was used for the analysis. Prior to both types of training there were no significant correlations between performance on either the deontic or arbitrary tasks and the AH4 test of Intelligence. This is inconsistent with the findings reported in Experiment 1 of this thesis and also diverges from the typical relationship between ability and performance on the arbitrary tasks reported by Stanovich. After obligation training there were no correlations between either of the deontic tasks and ability but there was between performance on the arbitrary tasks and ability (.53, $p<.01$ two-tailed). After rules training, performance on all the tasks was related to ability (.50, $p<.01$ (two-tailed) for obligation problem; .41 and .40, $p<.05$ (two-tailed) for permission and arbitrary problems respectively). In sum, after training on the obligation schema performance on the deontic problems was not related to intelligence however performance on the arbitrary problems was. In contrast, performance on both problem types was related to ability after rules training. This indicates that the training is more effective on these problems for individuals of higher ability.

	Obligation training AH4 total		Rules training AH4 total	
	Pre	Post	Pre	Post
Deontic	.27	.23	.27	.46*
Arbitrary	.32	.53**	.19	.40*

*p<.05, **p<.01, ***p<.001 (two-tailed)

Table 5.8 Correlations between reasoning tasks and ability

5.3.4 Thinking Styles

A correlational analysis was performed on the subscales of the rational-experiential inventory. As a moderate correlation was found between the Rational Ability and Rational Engagement subscales (.39, p<.01) and a highly significant correlation between the Experiential Ability and Experiential Engagement subscales (.74, p<.001), the four subscales were collapsed into two, the Rational and Experiential subscales respectively.

	Obligation training						Rules training					
	Rationality		Experiential		TDC		Rationality		Experiential		TDC	
	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post
deontic	0.23	0.32	-0.09	-0.18	0.47*	0.59**	0.26	0.37	-0.04	-0.16	-0.12	0.22
arbitrary	0.13	0.34	0.24	0.01	0.00	-0.09	0.11	0.32	0.09	0.02	0.12	0.25

*p<.05 **p<.01 ***p<.001 (two-tailed)

Table 5.9 Correlations between performance on the tasks pre and post training and thinking style scales under each condition

Scores on the rationality, experiential and Stanovich's thinking disposition composite subscales were correlated with scores on the AH/4 measure of intelligence and the only association found was between the rationality subscale of the REI and ability at .49 (p<.001). Of particular interest were the relationships between performance on each of the selection task problem types and the indices of rationality, experientiality and Stanovich's thinking disposition composite score. Table 5.9 displays the correlations under each condition.

In the obligation training condition there were no associations except for between the thinking disposition composite score and performance on the deontic type problems prior to training (.47, $p<.05$) and after training (.59, $p<.01$). In the rules condition there were no significant relationships between performance on any of the tasks and measures of thinking styles. In other words, performance on the deontic tasks was related to more open-minded, flexible thinking styles both before and after obligation schema training, but these relationships were not present under rules training. There were no significant relationships between performance on the arbitrary problems and thinking styles measures in either training condition.

5.3.5 Conditional reasoning

Of interest was whether either training had improved conditional reasoning scores. Rules training involved training on the material conditional therefore this training specifically should improve performance on these problems. Table 5.10 illustrates the total conditional reasoning scores and the mean number of logically correct inferences made relating to modus ponens, affirmation of the consequent, denial of the antecendent and modus tollens. There was some improvement after both types of training. A 2 x 2 x 3 (training condition x pretest/posttest x inferences) ANOVA was performed to observe whether training had an impact on any of the inferences. The modus ponens inference was not included in the analysis due to ceiling effects pre and post both types of training. There was no effect of training condition ($F(1, 58) = 0.16$, $MSE = 0.22$, $p>.1$) but there was a difference between pre and post scores overall ($F(1, 58) = 6.36$, $MSE = 2.02$, $p<.02$). The interaction between condition and the difference pre to post scores failed to reach significance ($F(1, 58) = 0.01$,

$MSE = 0.002$, $p>.1$). In other words there was a significant increase in logically correct responding on the conditional reasoning problems under both training conditions.

	Obligation training		Rules training	
	pre	post	pre	post
MP	2.00	2.00	2.00	1.97
AC	0.77	0.97	0.67	0.87
DA	0.73	0.87	0.60	0.80
MT	1.33	1.43	1.40	1.47
Total	4.83	5.27	4.67	5.10

Table 5.10 Mean inference rates for each problem type

A main effect of inference type was obtained ($F(2, 116) = 13.60$, $MSE = 15.76$, $p<.001$) and a Tukey HSD follow up revealed that participants achieved more logically correct responses on modus tollens inferences than either affirmation of the consequent or denial of the antecedent ($p<.001$ for each). No interaction between training condition and inference type was obtained ($F(2, 116) = 0.19$, $MSE = 0.22$, $p>.1$), and no interaction between inference type and pre to post training scores ($F(2, 116) = 0.26$, $MSE = 0.11$, $p>.1$). The three way interaction between training condition, inference type and pre to post training scores also failed to reach significance ($F(2, 116) = 0.05$, $MSE = 0.02$, $p>.1$). In summary it can be concluded that both types of training increase scores on all the inferences, apart from MP, which had a ceiling effect prior to training. Participants achieve more logically correct responses on the valid MP and MT inferences than the invalid AC and DA inferences.

5.4 Discussion - Experiment 4

In summary, the results of Experiment 4 demonstrated that obligation schema training facilitated reasoning performance on both arbitrary and deontic selection tasks, including problems designed to elicit both obligation and permission schemas. The training served to reduce p only responses on both the arbitrary and deontic tasks. Abstract training in rules of logic facilitated reasoning performance on arbitrary selection task problems only. No transfer occurred from the rules training to either of the deontic problem types. P/q responses were reduced on the arbitrary tasks under this training procedure. Both training procedures resulted in an increase in logical performance on the conditional reasoning problems, including blocking the invited inferences of AC and DA.

These findings are not quite as predicted for either rules training or obligation schema training. It was predicted that training in an abstract rule would transfer to arbitrary and deontic problems; however there was no transfer to obligation or permission schema problems, only to other arbitrary problems. According to dual process theory training in a domain-specific rule such as an obligation schema should transfer only to other obligation schema problems; however transfer occurred to obligation, permission and arbitrary problems therefore lending support to Cheng et al.'s original findings. The question is why are these conflicting results obtained?

Let's look at the rules training first. Cheng et al. reported transfer to all problem types after this training (apart from biconditionals). The main difference between their study and this one is that the training examples were changed from consisting of thematic content to arbitrary. It is possible then that by changing the examples, facilitation of transfer to deontic problems has disappeared altogether. This finding is in contrast with Klaczynski and Laipple (1993) who proposed that the rules are relatively general when giving abstract training therefore there is no concrete knowledge base on to which the rules can be mapped. Once the domain-general rule is induced, transfer occurs across different problem types. This view is consistent with dual process theories, which would argue that training in a domain-general rule should transfer across domains.

Obligation schema training on the other hand, did transfer to problems from other domains. This is inconsistent with the predictions discussed here which would predict transfer to problems from the same domain only. This finding does support the original findings of Cheng et al. who found transfer to two of their arbitrary problems although it was questionable whether their test problems were entirely abstract. The design of this experiment addressed this directly by ensuring the test problems used were completely arbitrary when compared to Cheng et al.'s. Examples were also used as part of the training procedure to make the training more equal in length and structure to the rules training which may also have had an influence.

This finding is in contrast to Klaczynski and Laipple (1993) who consistently found no transfer of training from Permission schema problems to arbitrary problems. They argued that when training assimilates to pre-existing domain-specific rules (as in permission or obligation schemas), the induction of a general rule is less likely to occur. They suggest that the induction of a domain-independent rule during the training occurs readily when training contradicts the problem-solvers' incorrect strategies for solving the problems and forces people to develop a new rule to accommodate the information presented. It could be that the obligation schema training procedure itself allows the induction of a domain-independent rule. Cheng et al. proposed that participants in their study were provided with information as to which cases constituted violations as well as orienting them toward checking violations, which may have facilitated performance on the abstract problems. As these are consistent with the material conditional they can also be applied to arbitrary problems.

So what do the correlations between performance on the selections tasks and ability tell us? Stanovich (1998a; 1999) proposed that higher ability participants are able to override the heuristic p/q response and reason analytically on the arbitrary selection task. However, no correlations were present prior to training with either arbitrary or deontic problems. After both types of training, associations were found with ability. After obligation training there was a significant correlation between performance on the arbitrary tasks and ability and after rules training, performance on both problem types was associated with ability. This may be best explained in dual process terms. System 2 is sensitive to instruction and training. It is also related to intelligence therefore if people are applying the rule they have been taught then these people may have more cognitive resources in which to do so. In other words, the correlation is an indicator of System 2 involvement in the task. The more resources they have available, the more effective the training.

Returning to the training effects found in this experiment, obligation training transferred to both deontic and arbitrary tasks. Therefore it seems that this training was more effective for higher ability participants when solving arbitrary problems. However, the training facilitated improved reasoning on deontic problems also which could indicate that the schema-based training was more easily transferable to further tasks designed to elicit a pragmatic reasoning schema, regardless of ability levels. This finding is inconsistent with those of Klaczynski and Laipple (1993) who claimed that intelligence is related to the ability to transfer between domain-specific schema problems. Only individuals of higher ability are able to identify the structural similarities between the problems.

The rules training transferred to arbitrary tasks only. However performance on both types of task was related to ability after training. Again if we look at the correlations as being an index of System 2 engagement then we can suggest that this training was most effective for performance on both types of problem for individuals with more cognitive resources available. This finding is again in contrast with Klaczynski and Laipple who proposed that only weak relationships with ability would be found when transferring from domain-independent problems as the rule may be more easily applied to target problems. It could be argued that it was the high ability participants in this experiment who firstly understood the training and were then able to apply it to arbitrary and deontic tasks. However, for individuals of lower ability the rules training may have been too difficult to understand resulting in no transfer at all, and in fact may even have confused them leading to changes from correct to incorrect responding on the deontic tasks.

On inspection of individual responses it appears that several participants do change from a predominantly incorrect response to a predominantly correct (p/not q) response after the rules training but this is offset by several participants who change from the correct response prior to training to an incorrect response. The mean ability score for those participants who changed to a correct response after training was 105.86 compared to a mean ability score of 99.33 for participants who changed from a correct to an incorrect response. This is a possible indication that the rules training only facilitates transfer across domains in participants of higher ability whereas participants of lower ability are perhaps confused or misunderstand the training therefore resulting in incorrect choices.

The findings discussed here are completely inconsistent with the findings related to Klaczynski et al.'s training studies, both in the direction of transfer after training and the relationships with ability. However Klaczynski et al.'s training was very different to Cheng et al.'s. It didn't involve any explanations of the rules or schema being used. The training involved presenting participants with problems followed by explanations of the correct and incorrect choices. The two types of training explored here so far, rules of the material conditional and obligation schema, resulted in the reduction of different types of responses on different problems. Rules training reduced the heuristic p/q response on arbitrary problems significantly whilst obligation training reduced p only responses on both arbitrary and deontic problems. That in itself illustrates the different nature of the training techniques. It appears that rules training engages an individual's analytic reasoning system but only those of higher ability and enables them to override the heuristically cued p q response on arbitrary tasks. Obligation training on the other hand, alerts participants as to which card falsifies or violates the rule on both arbitrary and deontic tasks (see Evans & Over, 1996). The higher ability participants in this experiment are able to transfer this knowledge to arbitrary tasks from deontic explanations.

In Experiment 5 the aim was to replicate the findings of Experiment 4 but in addition to compare Cheng et al.'s training to Klaczynski et al.'s in order to investigate whether the different findings could be explained by the training procedures themselves. A brief introduction to Experiment 5 will follow including further details of Klaczynski et al.'s training procedures.

5.5 Experiment 5

5.5.1 Introduction and rationale for Experiment 5

Klaczynski, Gelfand, and Reese (1989) conducted a series of experiments to understand the conditions under which reasoning transfers between problems on the selection task. They tested the effects of problem explanations and verbalization instructions on transfer from arbitrary or thematic problems to another set of arbitrary problems. Participants were given either five arbitrary problems or five thematic problems initially. In the explanation conditions, participants were read in-depth explanations of the correct and incorrect selections immediately after they completed each of the first five problems (see Chapter 2 for training examples). Klaczynski et al. found that transfer occurred from the initial arbitrary problems to a final set of arbitrary problems but no transfer occurred from thematic to arbitrary problems. The results were discussed in terms of a 'contextual similarity hypothesis'. That is, the more shared elements a problem has, the greater the similarity. Therefore, the more likely that reasoning on one problem will be transferred to the next problem.

An alternative explanation of the pattern of results was offered following their next experiment, using Cheng et al.'s (Cheng, Holyoak, Nisbett, and Oliver, 1986; Cheng and Holyoak, 1985) hypothesis that participants are able to formulate a general problem solving rule or schema from explanations on arbitrary problems which they can subsequently use when presented with further arbitrary problems. New arbitrary problems are mapped onto the pre-existing schema and are then solvable. Thematic problems are tied to experiences and contexts and any problem solving rule that is formed from explanations is therefore limited in generality. Ansburg and Shields (2003) proposed that the problem-specific feedback emphasised the surface structure of the permission problems rather than deep structure which explained the lack of transfer from the permission problems.

Due to the conflicting findings of Experiment 4 with Klaczynski et al. (Klaczynski, Gelfand, & Reese, 1989; Klaczynski & Laipple, 1993), the aim of Experiment 5 was to replicate the results of Experiment 4, whilst at the same time drawing direct comparisons with Klaczynski's training procedures. No changes were made to the training procedures presented in Experiment 4 but two training conditions were added, Permission schema training (Klaczynski and Laipple, 1993) and Abstract training (Klaczynski, Gelfand, & Reese, 1989). These two conditions were different from the Rules and Obligation training in that they didn't have training in abstract principles followed by the examples. Klaczynski et al.'s training consisted of four problems within a certain domain, permission or arbitrary, dependent on which condition the participant was in. These four source problems were each followed by an explanation of the correct and incorrect selections, which were read out to the participants. For each explanation, participants were first told the correct answer and were given the reasons for selecting both the p and not q cards.

Participants were also told why the two incorrect cards (not p and q) should also not be selected. The form of the training explanations was identical across the problem types (See Appendix A5.1 and A5.2 for the Permission and Abstract training). They reported transfer to other problem types from abstract training examples but only transfer to problems of the same type for permission schema training examples. These training procedures were not compared in the same study. By including them together plus the rules and obligation training, transfer effects may be investigated on the same set of test problems.

It was predicted that the findings from Experiment 4 would be replicated. Rules training will transfer to only arbitrary selection task problems. Obligation schema training will facilitate reasoning on selection tasks consisting of both arbitrary and deontic content. In terms of Klaczynski et al.'s training, it was predicted that Permission training would facilitate transfer to other problems that could be translated in terms of the permission schema only, whereas the Abstract training would allow for the induction of more general problem-solving rules therefore resulting in transfer to all types of selection task problems. We also included, once again, measures of ability to examine the extent to which ability mediates the effectiveness of the different training procedures.

5.6 Method

Design

The experiment consisted of a between-subjects design involving four conditions. Two of the conditions were the same as in Experiment 4, Rules plus examples training and Obligation schema training (Cheng et al., 1986). Two further conditions were added, Permission schema training (Klaczynski and Laipple, 1993) and Abstract training (Klaczynski, Gelfand & Reese, 1989). The dependent variables were the eight selection task problems from Experiment 3 plus two extra selection tasks designed to elicit the 'causal' schema (see Table 5.11).

Participants

One hundred and twenty participants were randomly allocated to one of the four training conditions (obligation group, rules group, permission group and abstract group). Total mean age of the sample was 20.65 years (st.d. = 4.96). All were students and postgraduates from the University of Plymouth but none had participated in any form of logic training before.

Materials

Rules plus Examples Training

30 participants were presented with the rules plus examples training (Cheng et al., 1986; exp. 1) used in Experiment 4.

Obligation Schema plus Examples Training

30 participants were given obligation schema training plus examples as used in Experiment 4.

Permission Schema Training

This training condition was adapted from Klaczynski and Laipple (1993, Experiment 2). Participants were given four permission problems to complete (if one is to take Action A, then one must first satisfy Precondition P), each followed by an explanation detailing the correct and incorrect responses. In Klaczynski and Laipple's study these explanations were read out to the participants immediately after each problem but for this experiment a written explanation was given (see Appendix A5.1).

For each explanation, participants were first told the correct answer and were given the reasons for selecting both the p and not q cards. In the final part of each explanation, participants were told why the two incorrect cards i.e. not p and q, should not be selected.

Because a common mistake on the four-card task involves selection of the q card and because selection of the not p card is rare, the rationale for not selecting q was particularly emphasised.

Abstract Training

This training procedure was the same as the Permission schema training but instead of problems designed to evoke the permission schema, arbitrary selection task problems were used. The form of the training explanations was identical across the two training problem types (permission and abstract), the only difference being the explanations given for the two types of problem were the specific instantiations of the four cards presented in the problems (p, q, not p, and not q). See Appendix A5.2 for the abstract training condition examples and explanations.

Cognitive ability

the AH4-Group test of General Intelligence was administered (Heim, 1967) as used in Experiments 1 and 4. The test consists of two parts, each containing 65 items. Part one is comprised of both verbal and numerical problems, part two is composed of problems in diagrammatic form. Each section is timed and must be completed in 10 minutes.

Selection task problems

As in Experiment 4, half of the problems were used as a pre-test and half as a post-test. The sixteen problems described in Experiment 4 were used plus four new ones which were designed to invoke the 'causal schema'. The new problems were adapted from Klaczynski and Laipple (1993). They contained rules describing a causal relationship between the antecedent and consequent (see Table 5.11). Presentation of the problems was balanced so that the problems that were presented as a pre-test for half the participants became the post-test for the other half and vice versa. T-tests performed on the two forms showed no difference between form 1 and 2 on the arbitrary problems at t(118) = 0.68, p=.50 but a significant difference between the two deontic forms at t(118) = -2.00, p<.05.

Form	Conditional statement	Response
Form 1	"If a person exercises frequently, then that person will be in good shape."	Exercises frequently (p) Rarely exercises (not p) Good shape (q) Poor shape (not q)
	"If a person eats excessively, then that person will gain weight."	Eats excessively (p) Always dieting (not p) Gained weight (q) Lost weight (not q)
Form 2	"If a patient has high cholesterol, then that patient will have high blood pressure"	High cholesterol (p) Low cholesterol (not p) High blood pressure (q) Low blood pressure (not q)
	"If a person drinks heavily, then that person will become intoxicated."	Ten pints of beer (p) Two lemonades (not p) Customer drunk (q) Customer not drunk (not q)

Table 5.11 Causal Conditional statements and their responses

Procedure

Participants were tested in groups of two to eight. Each session took between one hour and a quarter to one hour and a half to complete the study. Participants were each given a booklet which contained all the test materials and training to work through. Each participant was randomly allocated to one of the conditions. First they completed the two subtests of the AH4. They were then asked to complete the ten selection task problems that made up the pre-test part of the experiment before undergoing the training section. Participants were asked to read and complete the Training section carefully and to make sure they understood what they were being trained to do. When they reached the end of that section they were to wait for other members of the group to reach the same point before progressing on to the post-test selection tasks.

5.7 Results

5.7.1 Selection task responding under each training condition

The selection task problems were scored as the number correct within each type, two obligation, two permission, two causal and four arbitrary tasks. The mean proportion of correct responses for each type was then calculated and these scores were used for all analyses. Table 5.12 presents the mean proportion of correct responses for each type of selection task under each condition, pre and post training.

	Cheng's Training				Klaczynski's Training			
	Obligation training		Rules training		Permission training		Abstract training	
Problem type	pre	post	pre	post	pre	post	pre	post
obligation	.47	.70	.40	.38	.33	.72	.50	.77
permission	.63	.71	.45	.43	.53	.85	.65	.88
arbitrary	.05	.15	.03	.15	.04	.22	.02	.74
causal	.13	.27	.08	.30	.07	.27	.03	.65

Table 5.12 Mean proportion of correct responses for each type of selection task under each condition, pre and post training (N=30 per training condition).

Table 5.12 illustrates that prior to training mean scores on the obligation and permission tasks were higher than arbitrary or causal task scores. A repeated measures ANOVA performed on the pre-training scores collapsed across the training conditions show that there is a significant difference in correct responses dependent on problem type ($F(3, 357) = 107.35$, $MSE = 8.09$, $p<.001$). Further analyses provided evidence for Cheng et al.'s reasoning schema hypothesis and findings in the selection task literature as problems designed to evoke the permission schema yielded the highest number of correct responses compared to obligation, causal and arbitrary problems ($p < .001$ for each). Obligation problems yielded the next highest ($p< .001$ compared to arbitrary and causal) providing evidence for an obligation schema. Arbitrary and causal problem scores were not significantly different from each other ($p = .24$) as expected.

As both types of training in experiment 4 impacted on the problems designed to evoke obligation and permission schemas in the same way, for the following analyis they were collapsed into one variable, deontic problems. Performance scores on the causal problems were collapsed with the arbitrary tasks. The aim of the first analysis performed on the data was to investigate the effects of the two different types of schema-based training and the two different types of abstract training. This way we can examine the transfer effects of Klaczynski et al.'s training directly with Cheng et al.'s training.

A 2 x 2 between (deontic/abstract training x Cheng/Klaczynski training) x 2 within (problem type) ANOVA was performed on the data. There was no difference on scores overall dependent on whether the training was schema-based or abstract ($F(1,116) = 0.40$, $MSE = 0.07$, $p>.1$), however there was a difference dependent on whether the training was Cheng et al.'s or Klaczynski et al.'s ($F(1,116) = 10.81$, $MSE = 1.99$, $p<.01$). Participants in Klaczynski et al.'s training conditions performed better overall than in Cheng et al.'s training groups. The ANOVA yielded a significant interaction between the training and pre and post training scores which is illustrated in Figure 5.4 (($F(1,116) = 23.75$, $MSE = 2.12$, $p<.001$).

Figure 5.4 illustrates that prior to both types of training, performance on the selection tasks was the same. However, after training performance was improved under both conditions but more so after Klaczynski's training (($F(1,116) = 6.8$, $MSE = 0.6$, $p<.05$ for Cheng's; ($F(1,116) = 90.27$, $MSE = 8.07$, $p<.001$ for Klaczynski's). The difference between post-training scores under the two conditions was also significant ($F(1,116) = 23.41$, $MSE = 4.11$, $p<.001$). In other words, Klaczynski's training was far more effective at improving reasoning then Cheng's training. Due to the differences in Klaczynski and Cheng's training, for the following analysis they were investigated separately.

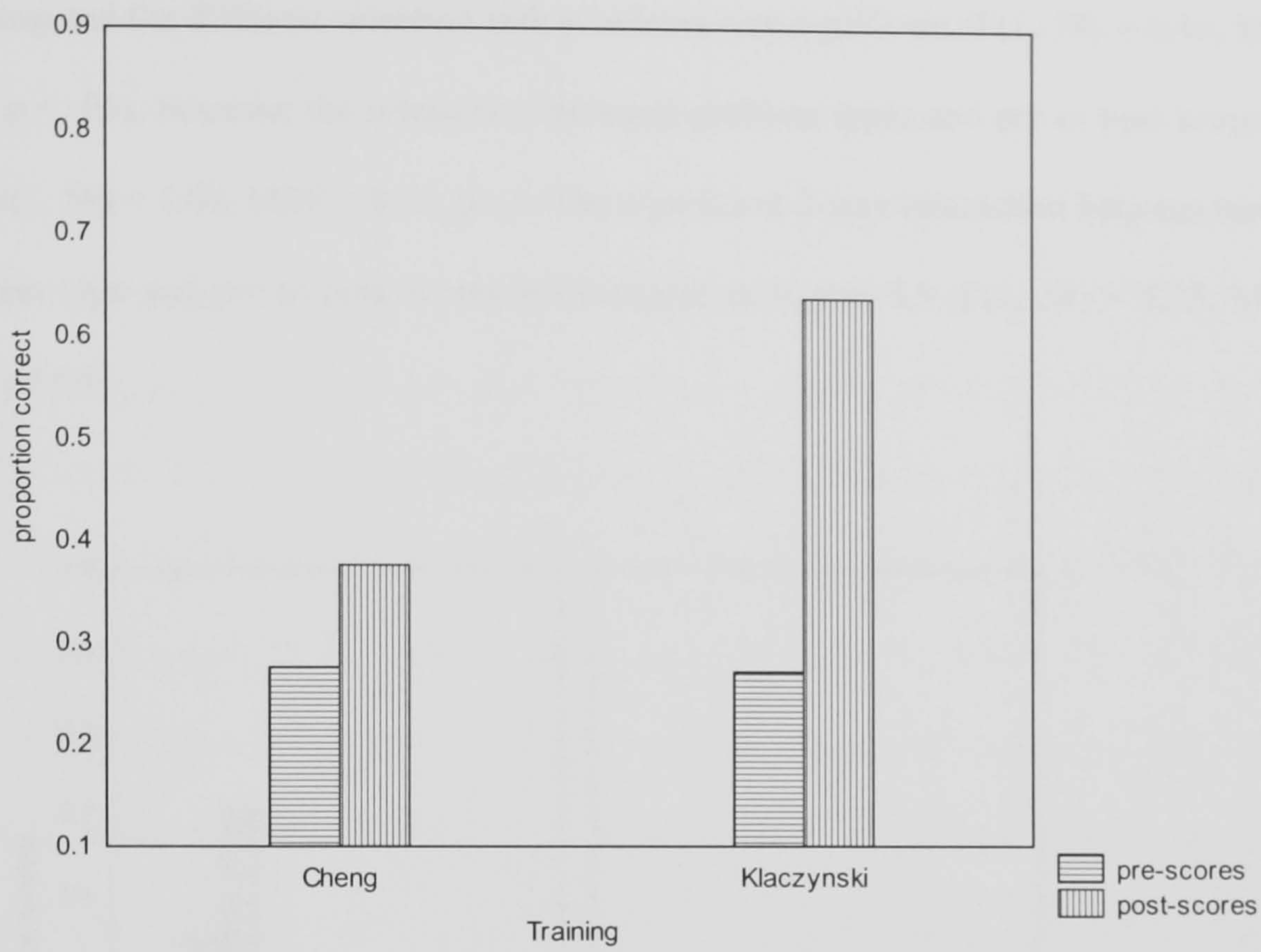

Figure 5.4 Interaction between Cheng and Klaczynski's training and performance scores pre and post training.

5.7.1.1 Cheng et al.'s obligation schema and rules training

First a 2 x 2 x 2 (Condition X Problem Type X Pre-Post Training scores) ANOVA was performed to observe any differences in correct responding (using proportion of correct scores) pre to post training on the selection tasks dependent on the Obligation and Rules training conditions. The difference between scores dependent on training just failed to reach significance ($F(1, 58) = 3.64$, $MSE = 0.73$, $p=.06$) with higher scores attained under the obligation training group, but there was a main effect of pre and post training scores ($F(1, 58) = 7.73$, $MSE = 0.61$, $p<0.01$) and problem type ($F(1, 58) = 95.25$, $MSE = 9.30$, $p< 0.001$) with deontic problems yielding the highest number of correct scores.

Of particular interest were the interactions between the problems, training and pre and post scores. A 2-way interaction between training and pre and post scores failed to reach significance ($F(1, 58) = 0.88$, $MSE = 0.07$, $p>.1$). Another 2-way interaction between Training and the different selection task problems was significant ($F(1, 58) = 6.40$, $MSE = 0.62$, $p < .05$), however the interaction between problem types and pre to post scores was not ($F(1, 58) = 1.64$, $MSE = 0.05$, $p>.1$. The significant 3-way interaction between training, problem type and pre to post scores is illustrated in Figure 5.5 ($F(1, 58) = 5.25$, $MSE = 0.17$, $p<.05$).

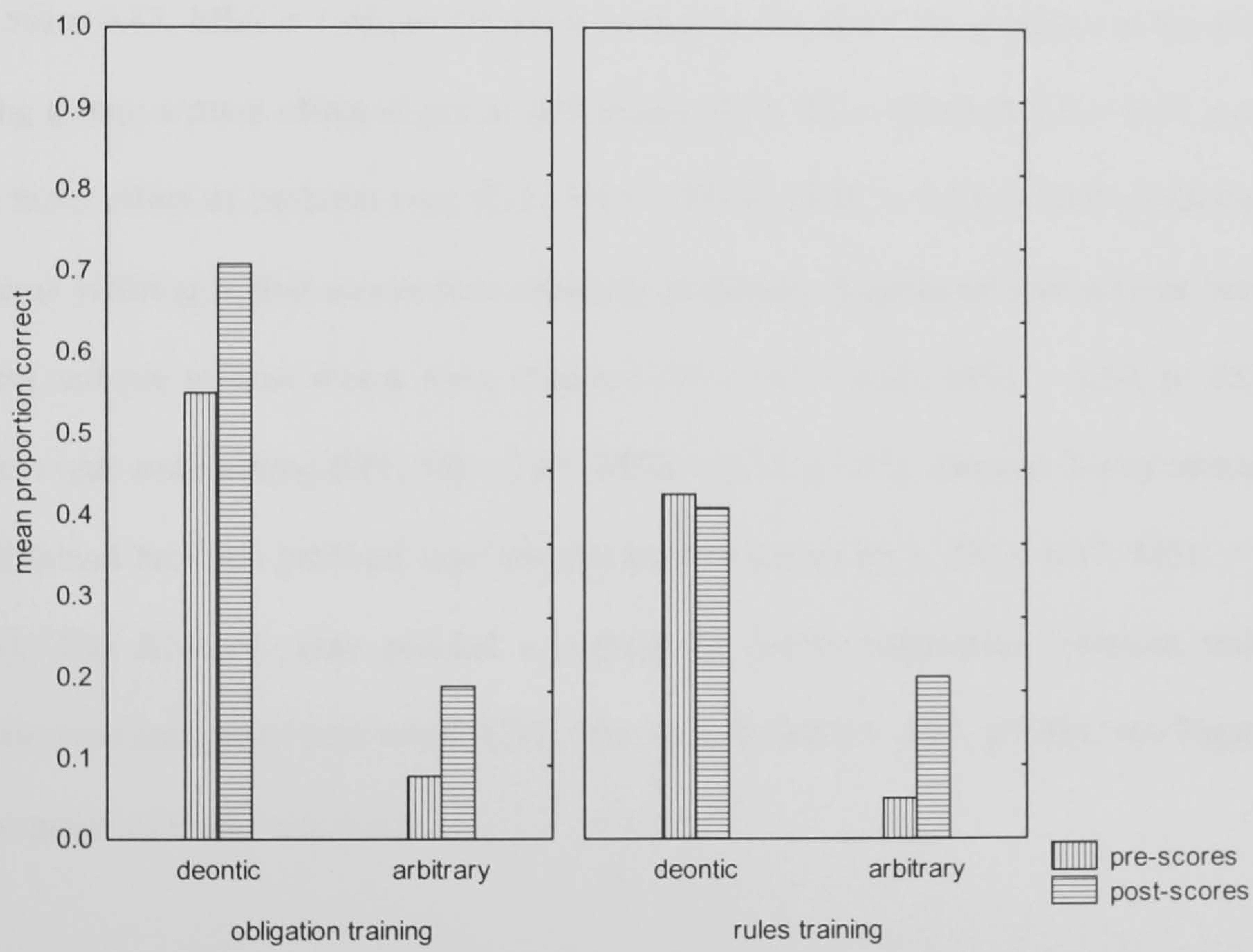

Figure 5.5 3-Way interaction between Training, Problem Types and pre and post training scores under Cheng et al.'s Obligation and Rules training.

There was an increase in scores after Obligation training on the deontic, ($F(1,58) = 5.77$, $MSE = 0.38$, $p<.05$) and arbitrary problems ($F(1,58) = 3.99$, $MSE = 0.18$, $p=.05$). After rules training there is only an increase on the arbitrary problems ($F(1, 58) = 7.29$, $MSE= 0.34$, $p<.01$). Rules training did not impact on the deontic problems ($F(1, 58) = 0.06$, MSE= 0.004, $p>.1$). In other words, there were transfer of training effects after obligation schema training on deontic problems and arbitrary problems, however performance after rules training was facilitated on arbitrary problems only. These findings replicate those of Experiment 4 although the impact on arbitrary problems after both types of training is not as strong.

5.7.1.2 Klaczynski's permission schema and abstract training

The same 2 x 2 x 2 (training x problem type x pre/post scores) ANOVA was performed on the Permission and Abstract training groups. The ANOVA yielded a main effect of training ($F(1, 58) = 9.13$, MSE = 1.53, $p<.01$) with scores on the tasks being higher in the abstract training group; a main effect of pre to post scores ($F(1, 58) = 80.69$, MSE = 8.07, $p<.001$; and a main effect of problem type ($F(1, 58) = 133.56$, MSE = 9.53, $p<.001$) with deontic problems yielding higher scores than arbitrary problems. Significant interactions between training and pre to post scores were obtained ($F(1, 58) = 6.00$, MSE = 0.60, $p<.05$) and problem type and training ($F(1, 58) = 3.89$, MSE = 0.28, $p=.05$). Another 2-way interaction was obtained between problem type and pre to post scores ($F(1, 58) = 6.47$, MSE = 0.27, $p<.05$). The ANOVA also yielded a significant 3-way interaction between training, problem type and pre to post scores ($F(1, 58) = 32.76$, MSE = 1.35, $p<.001$; see Figure 5.6 for the graph of the interaction).

On examination of the pre and post-training scores under Permission training for all the problem types it appears that transfer has occurred to all of them. Further analysis provides evidence in that the difference pre-post training scores is significant for the deontic tasks ($F(1, 58) = 22.55$, MSE = 1.84, $p<.001$) and the arbitrary tasks ($F(1, 58) = 8.44$, MSE = 0.50, $p< .01$). This contradicts the findings of Klaczynski, Gelfand, and Reese who found little transfer from thematic problems to abstract problems.

Finally, the largest effects appear to be under Abstract training. Prior to training, participants achieved higher scores on the deontic problems than arbitrary problems, but after Abstract training the differences between the scores on each problem type was decreased as well as performance enhanced for all problems. The difference in pre to post-training scores was significant for deontic problems ($F(1, 58) = 11.51$, $MSE = 0.94$, $p<.001$) and arbitrary problems ($F(1, 58) = 117.28$, $MSE = 7.00$, $p<.001$). This finding is consistent with Klaczynski, Gelfand and Reese who suggest the abstract training facilitates the formation of a more general problem-solving schema which may be used to solve subsequent problems. In summary, permission and abstract training improved performance on the selection tasks within the same domain of training the most, suggesting a degree of domain specificity, but also transferred to problems in different domains.

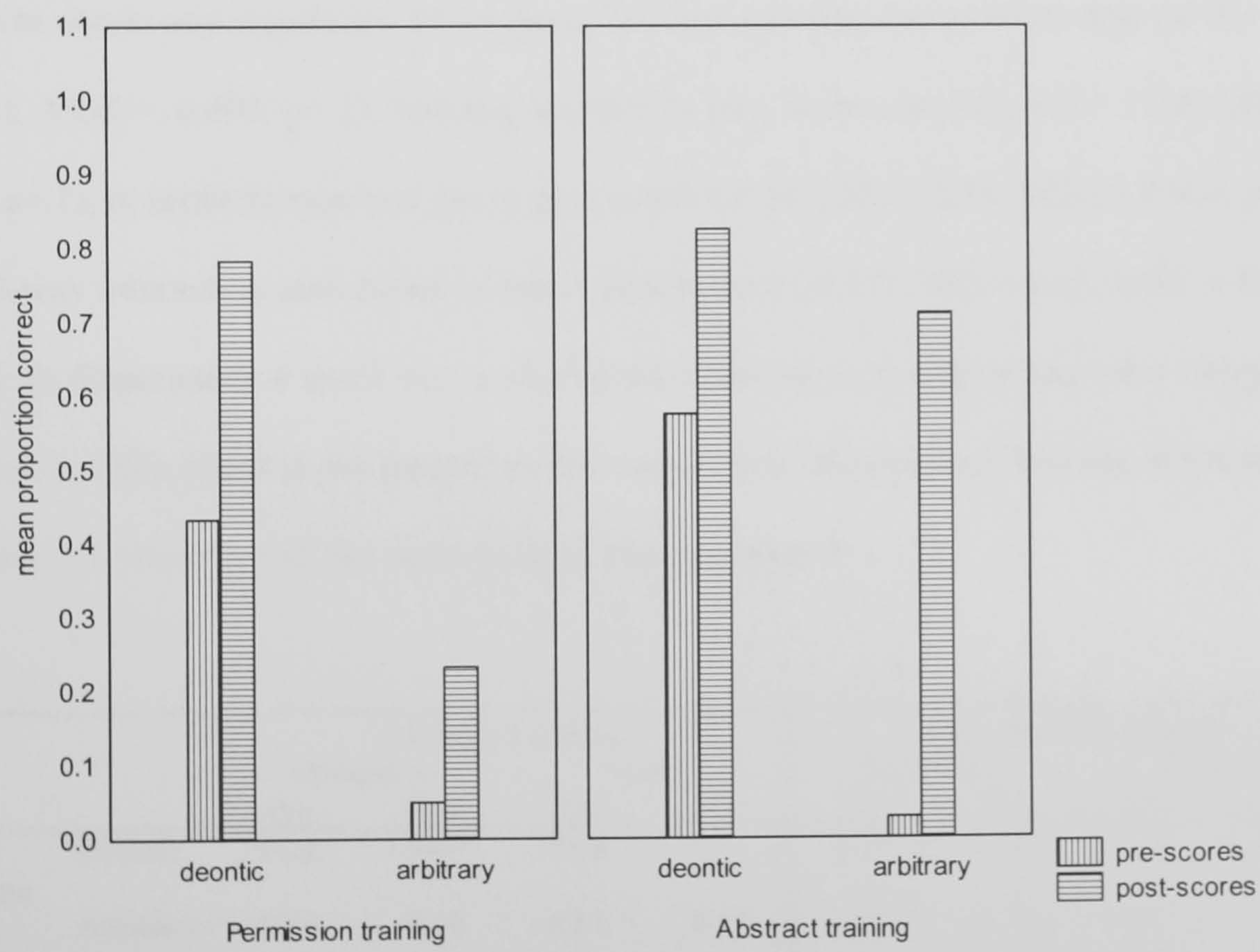

Figure 5.6 3-Way interaction between Training, Problem Types and pre and post training scores under Klaczynski's Permission schema and Abstract training.

5.7.2 Changes in response patterns

As in experiment 4 an investigation into changes in response patterns dependent on the training condition was conducted. Again, due to the differences in training, the obligation and rules training was looked at separately to the permission and abstract training. Table 5.13 illustrates p only and p/q responses under each training condition.

5.7.2.1 Cheng et al.'s obligation schema and rules training

A 2 x 2 x 2 (training x problem type x pre/post scores) ANOVA was conducted on p only responses under obligation and rules training. No effect of training was obtained (at $F(1, 58) = 0.01$, $MSE = 0.002$, $p>.1$), and no effect of pre to post training scores (at $F(1, 58) = 0.20$, $MSE = 0.02$, $p>.1$). There was an effect of problem type (at $F(1, 58) = 7.81$, $MSE = 0.27$, $p<.01$) with more p only responding on the arbitrary tasks overall. The ANOVA also failed to obtain any significant interactions between training and problem type (at $F(1, 58) = 0.01$, $MSE = 0.000$, $p>.1$), training and pre to post scores (at $F(1, 58) = 0.68$, $MSE = 0.06$, $p>.1$), or problem type and pre to post scores (at $F(1, 58) = 0.34$, $MSE = 0.000$, $p>.1$). The 3-way interaction also failed to reach significance (at $F(1, 58) = 0.61$, $MSE = 0.007$, $p>.1$). In Experiment 4 there was a significant decrease in p responding after obligation training but this effect is not present in this experiment. However on looking at the scores in Table 5.13 there is still the same trend in this experiment.

		Cheng's Training				Klaczynski's Training			
		Obligation		Rules		Permission		Abstract	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post
P response	Deontic	0.12	0.07	0.08	0.12	0.20	0.06	0.02	0.04
	Arbitrary	0.18	0.14	0.17	0.16	0.19	0.11	0.09	0.03
P q response	Deontic	0.22	0.17	0.28	0.24	0.22	0.10	0.26	0.06
	arbitary	0.54	0.50	0.59	0.36	0.66	0.45	0.73	0.11

Table 5.13 Mean p and p/q responses before and after each training condition.

The same 2 x 2 x 2 ANOVA was performed but looking at p/q responding before and after Obligation and Rules training. There was no main effect of training (at $F(1, 58) = 0.03$, $MSE = 0.01$, $p>.1$) but there was a main effect of problem type (at $F(1, 58) = 49.53$, $MSE = 4.40$, $p<.001$) with higher rates of p/q responding on the arbitrary tasks. The ANOVA also yielded a main effect of pre to post scores (at $F(1, 58) = 5.17$, $MSE = 0.53$, $p<.05$) with a decrease in p/q responding after training.

There were no significant interactions between training and problem type (at $F(1, 58) = 2.08$, $MSE = 0.18$, $p>.1$), training and pre to post scores (at $F(1, 58) = 1.07$, $MSE = 0.11$, $p>.1$), or problem type and pre to post scores (at $F(1, 58) = 3.12$, $MSE = 0.12$, $p>.05$). The ANOVA did yield a significant 3-way interaction between training, problem type and pre to post scores however ($F(1, 58) = 4.17$, $MSE = 0.16$, $p<.05$. See Figure 5.7 for the graph of the interaction).

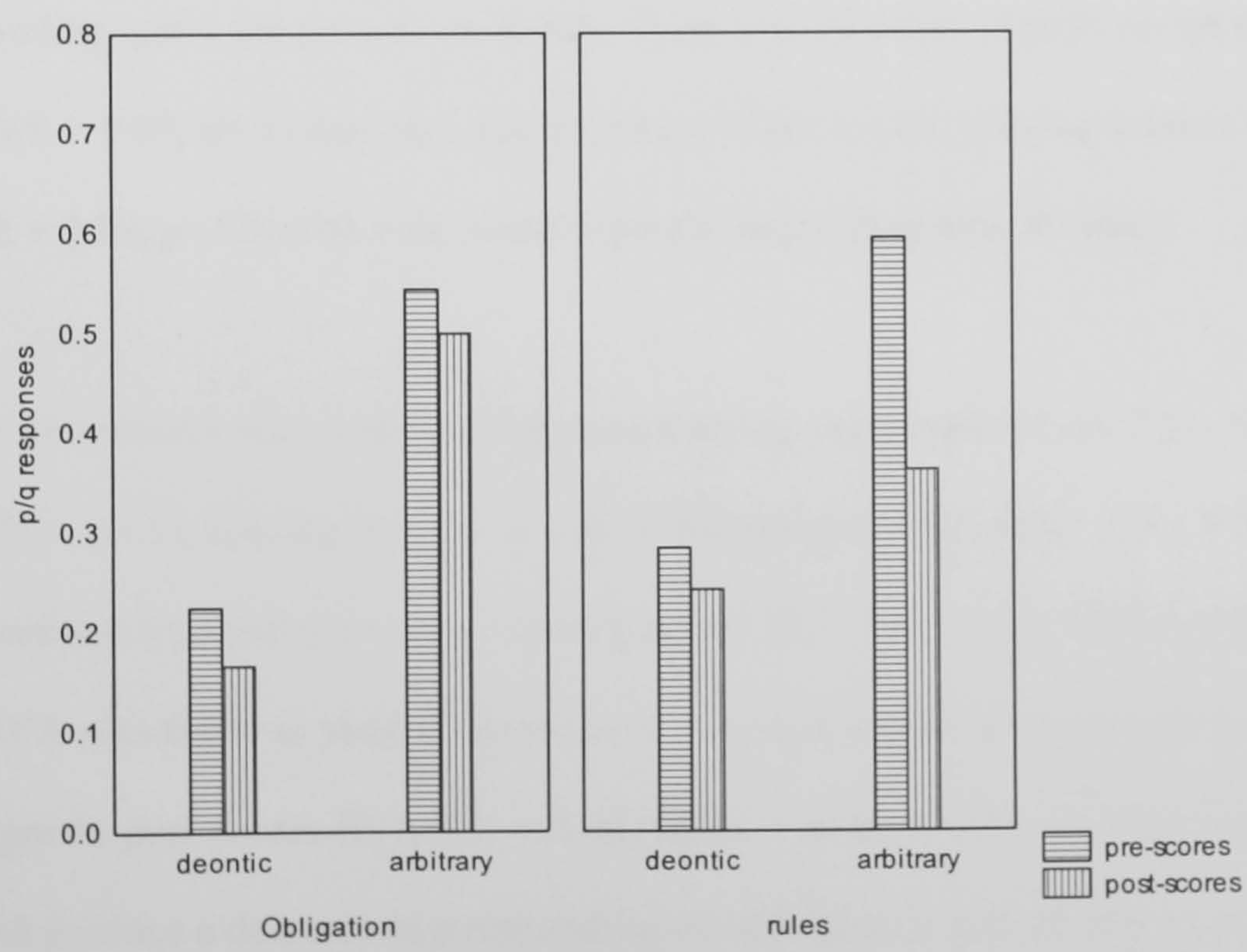

Figure 5.7 3-Way Interaction between Training, Problem Types and pre and post Training p/q scores under Cheng et al.'s Obligation and Rules training.

Figure 5.7 illustrates a drop in p/q responding on both deontic and arbitrary problems after obligation and rules training. A series of planned comparisons revealed that under obligation training the decrease in p/q responding was not significant for either deontic tasks ($F(1, 58) = 0.82$, $MSE = 0.05$, $p>.1$) or arbitrary tasks ($F(1, 58) = 0.37$, $MSE = 0.03$, $p>.1$). After rules training the decrease in p/q responding on the deontic problems was not significant ($F(1, 58) = 0.42$, $MSE = 0.03$, $p>.1$) however the decrease on the arbitrary problems was ($F(1, 58) = 10.30$, $MSE = 0.82$, $p<0.01$). In summary, the findings in this experiment replicate those of Experiment 4. P/q responding on the arbitrary tasks was significantly reduced after training in the material conditional rule system.

5.7.2.2 Klaczynski's permission schema and abstract training

A 2 x 2 x 2 (training x problem type x pre/post p responding scores) ANOVA was performed on the data from participants in the permission and abstract training groups. A main effect of training was obtained $F(1, 58) = 4.87$, $MSE = 0.53$, $p<.05$) with more p only responding under the permission group. There was no effect of problem type ($F(1, 58) = 1.16$, $MSE = 0.03$, $p>.1$) but there was an effect of pre to post training scores ($F(1, 58) = 4.02$, $MSE = 0.26$, $p<.05$) with a decrease in p only responding after training.

No 2-way interactions were obtained between training and problem type ($F(1, 58) = 0.07$, $MSE = 0.002$, $p>.1$), training and pre to post training scores ($F(1, 58) = 1.98$, $MSE = 0.13$, $p>.1$) or problem type and pre to post training scores ($F(1, 58) = 0.17$, $MSE = 0.003$, $p>.1$). The ANOVA also failed to yield a significant 3-way interaction between training, problem type and pre to post scores ($F(1, 58) = 3.05$, $MSE = 0.05$, $p>.05$). In sum, Klaczynski's training did produce a decrease in p responding on both deontic and arbitrary tasks.

Another 2 x 2 x 2 (training x problem type x pre/post p/q responding) ANOVA was performed on the data under the permission and abstract training groups. There was no effect of training (F(1, 58) = 2.23, MSE = 0.31, p>.1) but there was a main effect of problem type (F(1, 58) = 91.09, MSE = 6.36, p<.001) and pre to post training scores on p/q responding (F(1, 58) = 47.75, MSE = 5.03, p<.001). In other words, arbitrary problems yield a higher number of p/q responses overall and training decreases the amount of p/q responding.

The 2-way interaction between training and problem type was marginally significant (F(1, 58) = 3.90, MSE = 0.27, p=.05) and the interactions between training and pre to post scores and problem type and pre to post scores were also significant (F(1, 58) = 8.41, MSE = 0.89, p<.01 and F(1, 58) = 22.75, MSE = 0.97, p<.001 respectively). There was also a 3-way interaction between training, problem type and pre to post scores (F(1, 58) = 9.95, MSE = 0.42, p<.01).

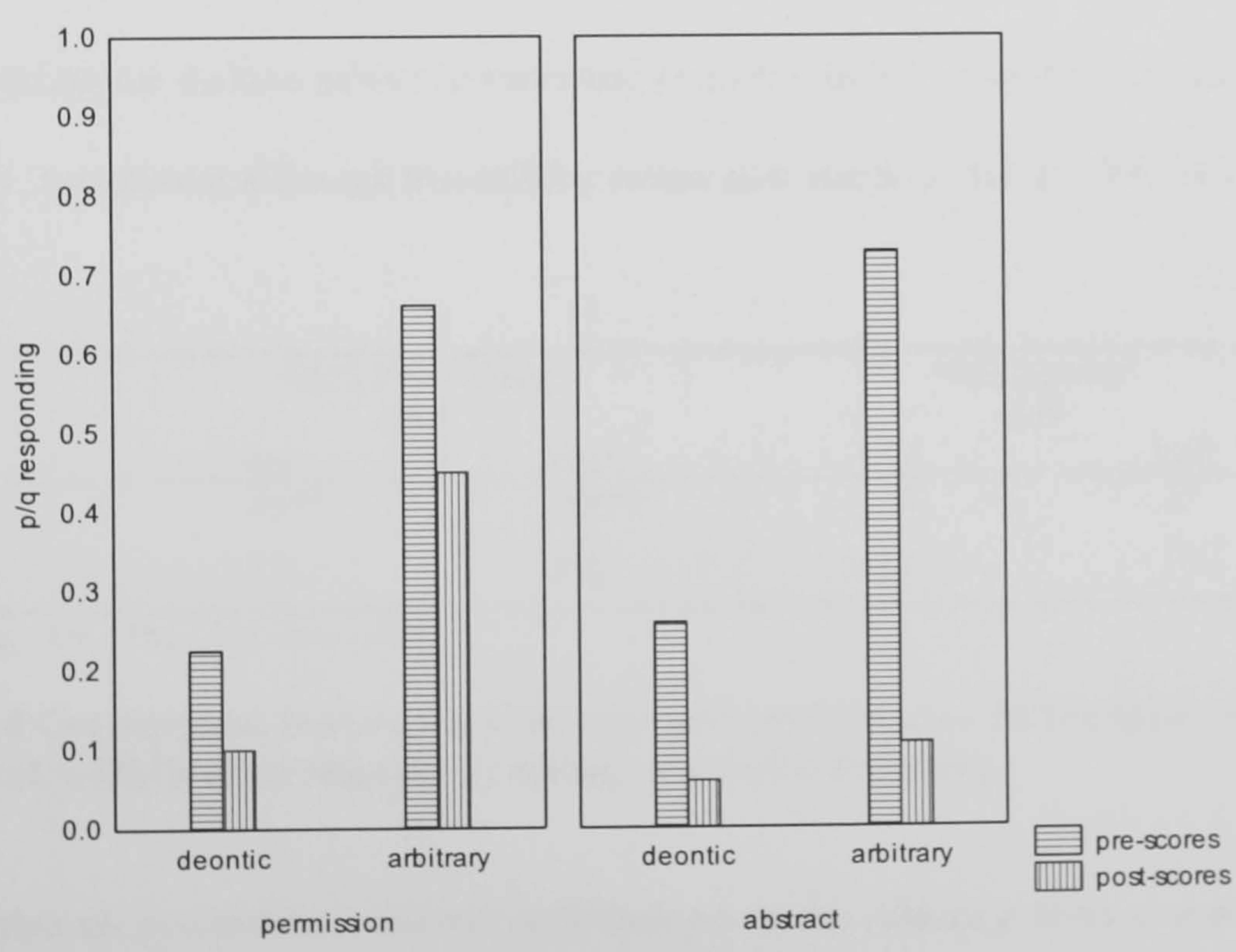

Figure 5.8 3-Way Interaction between Training, Problem Types and Pre and Post Training p/q Scores under Klaczynski's Permission and Abstract training

Figure 5.8 illustrates the 3-way interaction. A series of planned comparisons revealed that under permission training, p/q responding decreases significantly on arbitrary tasks ($F(1, 58) = 7.94$, MSE = 0.67, $p<.01$) but fails to reduce significantly on deontic tasks ($F(1, 58) = 3.68$, MSE = 0.23, $p<.05$). Under abstract training, p/q responding decreases on both deontic and arbitrary tasks ($F(1, 58) = 9.41$, MSE = 0.60, $p<.01$ and $F(1, 58) = 68.95$, MSE = 5.81, $p<.001$ respectively). The interaction is caused by the large reduction of p/q responses on arbitrary problems under abstract training.

5.7.3 Cognitive ability and performance under Cheng et al.'s Obligation and Rules training conditions

Table 5.14 shows the correlations between performance on the selection tasks and cognitive ability under Obligation schema training and Rules training. Looking at the correlations under obligation and rules training initially, it appears that the results are inconsistent with the findings of Experiment 4. There is now a correlation between performance on the deontic selection tasks and cognitive ability prior to obligation training (.49, $p<.01$, two-tailed) although this still increases post training (.66, $p<.001$, two-tailed).

	Obligation training AH4		Rules training AH4	
	Pre	Post	Pre	Post
Deontic	.49**	.66***	.36	.17
Arbitrary	.13	.26	.11	.12

*p< .05, **p< .01, ***p<.001 (two-tailed)

Table 5.14 Correlations between intelligence and performance on the tasks under Cheng et al.'s Obligation Schema Training and Rules Training

There is also no association between performance on the arbitrary tasks and ability after obligation training which was found previously. Under rules training the pattern of correlations after training had also disappeared indicating no associations with ability either before or after training on either type of task. One possible explanation for these findings is that the groups of participants in Experiments 4 and 5 differ in ability levels. As illustrated in Table 5.15 the mean ability scores under Obligation and Rules training in Experiments 4 and 5 do differ with lower scores being attained in Experiment 5. An independent t-test was performed which showed a significant difference between the two groups ($t(118) = 3.60$, $p<.001$) with higher ability participants in Experiment 4. We will return to this finding in the next section.

Training	AH4 TOTAL MEAN	
	Experiment 4	Experiment 5
Obligation	101.10	92.77
Rules	102.00	93.10

Table 5.15. Mean AH4 scores under Obligation and Rules training for Experiments 4 and 5.

5.7.4 Comparisons between performance in Experiments 4 and 5 under Cheng et al.'s Rules and Obligation training

Due to the differences in ability levels between participants in the two experiments a further analysis was conducted to investigate whether training was more effective in Experiment 4 compared to Experiment 5. A 2 x 2 x 2 x 2 (Experiment x Training x pre/post scores x problem type) ANOVA was performed. The main effect of Experiment was approaching significance ($F(1, 116) = 3.26$, $MSE = 0.77$, $p=.07$) illustrating that higher reasoning scores were obtained in Experiment 4 overall. The other finding of interest was a highly significant interaction between training, problem type and pre/post scores ($F(1, 116) = 14.32$, $MSE = 0.53$, $p<.001$). This did not interact with Experiment suggesting that the 3-way interaction is a robust effect ($F(1, 116) = 0.54$, $MSE = 0.02$, $p=.46$).

In summary, the main effect of experiment suggests that participants in Experiment 4 performed generally better than in Experiment 5. This is consistent with the AH4 data and suggests ability differences between the groups and less System 2 involvement amongst the participants in Experiment 5. This would also explain the absence of correlations between ability and performance in Experiment 5. The presence of a correlation between performance on the deontic problems and ability prior to training is consistent with Newstead, Handley, Harley, Wright, and Farrelly (2004) with a low ability group. The fact that the participants in this experiment are lower in ability than in Experiment 4, plus the absence of the correlations that were found previously and the smaller training effects leads us to the conclusion that participants in Experiment 5 were less likely to acquire the rule and thus less likely to apply it on further selection tasks. If ability is a marker of System 2 involvement then training in this experiment is not as effective in engaging System 2's analytic reasoning processes.

5.7.5 Cognitive ability and performance under Klaczynski's Permission and Abstract Training conditions

Table 5.16 illustrates the correlations between performance on the reasoning tasks and ability under Permission training and Abstract training. Under the Permission training group there was a moderate correlation between performance on the deontic tasks prior to training and ability (.45, p<.05, two-tailed) but after training this increased (.70, p<.001, two-tailed). A similar but weaker pattern was observed with arbitrary tasks also (.43, p<.05, two-tailed). Under the Abstract training group there were no significant correlations prior to training but after performance on both deontic and arbitrary tasks there was an association with intelligence (.43 and .47 respectively, both p<.05, two-tailed). Using ability as an indicator of System 2 involvement we may conclude that both types of training effectively engage analytic reasoning on deontic and arbitrary tasks in individuals with higher ability.

	Permission training AH4		Abstract training AH4	
	Pre	Post	Pre	Post
Deontic	.45*	.70***	.12	.43*
Arbitrary	.33	.43*	.10	.47*

*p<.05, **p<.01, ***p<.001 (two-tailed)

Table 5.16 Correlations between intelligence and performance on the tasks under Klaczynski's Permission training and Abstract training

5.8 Discussion - Experiment 5

Consistent with the findings in Experiment 4, Obligation schema training facilitated reasoning on both deontic and arbitrary problems, and Rules training facilitated reasoning on arbitrary problems only. The comparison between Cheng et al.'s training and Klaczynski's demonstrated that Klaczynski's training procedures were far more effective in improving reasoning responses. Reasoning performance was increased dramatically on both deontic and arbitrary selection tasks after both of Klaczynski's training procedures, although the effects were greater to problems within the same domain of training than across domains. In other words, Permission training increased correct responding more on further deontic tasks than arbitrary, and Abstract training increased correct responding more on the arbitrary tasks than deontic.

The replication of the transfer effects after Obligation and Rules training in this experiment, casts doubt on the original work of Cheng et al. Cheng et al. claimed that the Obligation training transferred to obligation schema problems and arbitrary problems that may be interpreted as obligations. The abstract test items they used in their study were thematic and it was thought that they might have affected the transfer. However, in these two experiments we have demonstrated that training in obligations transfers to purely arbitrary problems. The items have been used before in previous experiments and have been shown to elicit patterns of responses consistent with abstract selection tasks. The addition of two obligation schema examples to the training procedure may have caused these transfer effects however further discussion of this will be saved for the general discussion at the end of this chapter.

The findings in relation to Rules training are inconsistent with the findings of Cheng et al. They found that the training improved reasoning on a variety of problem types; however in neither of the experiments reported here did Rules training facilitate reasoning on any other problems other than ones consisting of arbitrary content. The most likely explanation for this lack of transfer is that the examples used as part of the training in the two experiments reported here were changed from consisting of thematic content to arbitrary. We propose that the thematic examples employed by Cheng et al. facilitated the transfer to other problems involving thematic content. The question is why the abstract training only transferred to abstract problems. It was predicted that abstract rules training would transfer to both arbitrary and deontic problems. Further discussion of this will be presented in the general discussion.

Overall the smaller facilitation effects resulting from these two training conditions in Experiment 5 lead us to conclude that the training procedures were less effective. The absence of the correlations with ability that were present in Experiment 4 along with the finding that the participants were of lower ability indicate that the training was less effective because participants did not have the cognitive resources available to acquire and apply the rules being taught. Also in comparing the experiments, performance overall in Experiment 5 was worse than Experiment 4 which is consistent with the notion that training facilitates System 2 thinking and is generally less effective the less resources are available.

Klaczynski et al.'s Abstract training facilitated performance on all problem types as expected. This contrasts with the effects found after Cheng et al.'s training. P only and p/q responses were significantly reduced on all tasks after training. In other words, the matching bias response was inhibited by the training procedure. Permission training on the other hand also resulted in improved performance on both deontic and arbitrary problems which was inconsistent with Klaczynski and Laipple. The training reduced p responding on both problem types however p/q responding was reduced for arbitrary problems only. This could be partly due to the low number of p/q responses on the deontic tasks prior to training in the first place. Overall, the larger effects of training to problems within the same domain as training i.e. from permission to deontic and from abstract to arbitrary, suggests some domain-dependence of training.

The pattern of correlations with ability after training was also inconsistent with Klaczynski et al.'s findings. Klaczynski and Laipple found that transferring from an abstract rule was not related to ability but transferring from a permission schema was. In Experiment 5, ability was related to both problem types after both training conditions. Again this can be viewed as an indicator of System 2 involvement on the tasks, with some increased involvement after training.

Interestingly, the absence of correlations between performance on arbitrary tasks and ability in both Experiments 4 and 5 prior to training is inconsistent with the findings reported in Chapter 3 (Experiment 1) and those reported by Stanovich (1998a; 1998b; 1999). Stanovich consistently finds associations between correct responding on the task and cognitive ability. He proposes that it is the higher ability participants who are able to decontextualise the problem from the items named in the rule (the heuristic p/q response); therefore resulting in the correct response.

Why is there such a big difference in training effects between Cheng et al.'s training and Klaczynski et al.'s training? Cheng et al.'s training procedures both involve lengthy explanations of the rule that participants are to learn prior to the presentation of examples. Klaczynski et al.'s training on the other hand involves the presentation of examples followed by explanations of how to solve the problems. In sum this procedure is a great deal simpler to follow and does not require participants to apply a set of learnt rules. Participants are provided with the strategies that they can then utilise on subsequent selection tasks, regardless of content. Again further discussion of this will be saved for the general discussion.

In Experiment 6 we will attempt to replicate the findings in relation to Klaczysnki et al.'s training. Klaczynski and Laipple found no transfer of training to problems in other domains than the permission problems after permission schema training. It is possible that the pretest/posttest design that is used here may have cued participants' responses. Prior to training in Experiment 5 participants completed both arbitrary and deontic selection task problems. Therefore when the participants were presented with either Abstract or Permission training, they were already aware of content variation within the tasks. The pretest was hence not used for this experiment. Another possible reason for conflicting results was order of presentation of the posttest selection tasks. In Experiment 5, the selection tasks were presented in random order, which may have affected individuals' responses in some way. That is receiving a mixed order of abstract and deontic tasks may have facilitated the recognition of the similarities between the rules in terms of underlying structure. Therefore to ensure that this did not affect reasoning performance, in the following experiment presentation of the tasks will be controlled.

With these effects controlled for, it will be possible to test Klaczynski's hypothesis that the Permission training will facilitate transfer to other problems that may be translated in terms of the permission schema only against the findings of Experiment 5 that permission schema training results in positive transfer to problems in other domains.

5.9 Aims for Experiment 6

The aim of Experiment 6 was to follow up the findings of Experiment 5 in relation to Permission training, whilst controlling the order of problem presentation and omitting the pre-test. Performance after Permission training was compared to a control group. Within each condition there were three groups therefore resulting in six groups altogether which differed in order of test items. Group One were presented with four arbitrary selection tasks followed by the permission problems. Group Two were presented with four permission problems followed by the arbitrary problems and Group Three were presented with all eight selection task problems in random order.

With these experimental variables controlled for, it would be possible to explore whether the Permission training would facilitate transfer to other problems that could be translated in terms of the permission schema and problems involving arbitrary content or, in accordance with dual process theories, Permission training would transfer to only problems consisting of deontic content.

5.10 Method

Design

The experiment involved a between subjects design with two conditions. The training condition was the same as used in Experiment 5, Klaczynski and Laipple's (1993) Permission training. The second condition was a Control group who received no training.

The dependent variables were eight selection task problems, four designed to elicit the permission schema and four arbitrary tasks.

Participants

One hundred and forty four participants were randomly allocated to one of the six conditions resulting in 24 participants per experimental group. Total mean age of the sample was 20.35 st.dev. 4.91 (control = 19.53, st.dev. 3.29; training = 21.18, st.dev. 6.03).

All were 1st year Psychology undergraduates from the University of Plymouth participating as part of a points scheme. None had taken part in any form of logic training beforehand.

Materials

Permission Training

Klaczynski's Permission training was used as in Experiment 5.

Selection Task Problems

Four problems designed to elicit the permission schema and four arbitrary selection tasks were taken from Experiments 4 and 5 (see Table 5.17).

Tasks	Conditional statement	Response
Arbitrary	"If the letter A is on one side of a card, then the number 7 must be on the other side of the card."	A (p) D (not p) 7 (q) 3 (not q)
	"If Triangle is on one side of a card, then Dog must be on the other side of the card."	Triangle (p) square (not p) dog (q) cat (not q)
	"If Yellow is on one side of a card, then Oak must be on the other side."	yellow (p) green (not p) oak (q) beech (not q)
	"If Coffee is on one side of a card, then Goldfish must be on the other side of the card."	coffee (p) tea (not p) goldfish (q) hamster (not q)
Permission	"If a person is travelling to the United States of America, then that person must have a visa."	United States of America (p) France (not p) Passport and a Visa (q) Passport and no Visa (not q)
	"If a person is drinking beer, then that person must be at least 18 years old."	Drinking beer (p) Drinking Coke (not p) 18 years of age (q) 16 years of age (not q)
	"If a student is withdrawing books from the library, then they must have a valid Identity card"	Books on loan (p) Books not on loan (not p) Valid Identity card (q) Invalid Identity card (not q)
	"If a person purchases cigarettes or tabacco, then they must be 16 years or over in age."	Bought cigarettes (p) Didn't buy cigarettes (not p) 17 years of age (q) 15 years of age (not q)

Table 5.17. Arbitrary and Permission Problems used in Experiment 6.

Cognitive Ability

The AH4-Group test of General Intelligence was administered (Heim, 1967) as used in Experiment 1, 4 and 5. The test consists of two parts, each containing 65 items. Each section is timed and must be completed in 10 minutes.

Procedure

Participants were tested in groups of two to eight. Testing sessions took approximately 30 minutes for the control Group and 45 minutes for the permission training group. Participants under the training condition worked through their training booklet first. They were instructed to read and complete the training section carefully and to make sure they understood the explanations of the correct responses presented after each problem. They then completed the two subtests of the AH4 before being presented with the test booklet containing the selection task problems. Participants who received no training completed the AH4 subtests followed by the selection tasks.

On completion of the study, participants were thanked and given an explanation of the study they had just taken part in.

5.11 Results – Experiment 6

5.11.1 Training effects on deontic and arbitrary selection task performance

The selection task problems were scored as the number correct within each type, four arbitrary problems and four deontic tasks designed to elicit the permission schema. Table 5.18 presents the proportion of correct responses for each type of selection task under each condition, the control group and Permission training group. It can be seen that arbitrary tasks yield the lowest number of correct responses compared to deontic tasks as expected.

Participants in the training condition achieved higher rates of correct responding on both types of task but further analysis is required to observe whether this is significant.

	Control		Permission Training	
	Mean	St.Dev	Mean	St.Dev
Arbitrary	.03	.14	.12	.29
Deontic	.70	.36	.81	.32

Table 5.18 Proportion of correct responses for each type of selection task under each condition, Control and Permission training (N=72 per condition)

A 2 Condition (control or training) x 3 Order (task presentation) x 2 Problem Type (arbitrary or deontic) ANOVA was performed on the data to observe any significant differences. A main effect of condition (F(1, 138) = 7.87, MSE = 0.68, p<.01) showed that training increases the amount of correct responding on the selection tasks. No effect of order was found (F(2,138) = 1.18, MSE = 0.10, p>.1), therefore it appears that there are no carry over effects between different types of tasks. The interaction between Condition and Order of tasks did also not reach significance (F(2,138) = .379, MSE = 0.03, p>.1) which illustrates that regardless of order, responses increased the same for all the participants under the training condition.

A main effect of Problem Type was found, (F(1,138) = 402.35, MSE = 33.69, p<.001), with Deontic tasks yielding much higher number correct scores than Arbitrary tasks as expected. There were no significant interactions between Problem type and Condition (F(1, 138) = 0.09, MSE = 0.01, p>.1, which would indicate that Permission schema training increases correct responding on Permission problems but also transfers to arbitrary problems as found in experiment 5 (see Figure 6). There were no significant interactions between Problem type and Order (F(2, 138) = 0.25, MSE = 0.02, p>.1), or Problem type, Condition and Order (F(2, 138) = 0.49, MSE = 0.04, p>.1).

5.11.2 Differences in response

There was no pre-test in this experiment therefore responses after training were compared to participants in the control condition who received no training (see Table 5.19 for p only and p/q responses under each condition). A 2 x 2 (condition x problem type) ANOVA performed on p only responses yielded a significant main effect of condition ($F(1, 142) = 13.53$, $MSE=15.12$, $p<.001$) with p only responses being lower after Permission training. There was also a significant effect of problem type ($F(1, 142) = 13.80$, $MSE=8.68$, $p<.001$) illustrating more p responses on arbitrary problems than deontic. The ANOVA also yielded a significant interaction ($F(1, 142) = 6.38$, $MSE = 4.01$, $p<.01$). Planned comparisons revealed that in the control condition p only responses were much higher for the arbitrary tasks than the deontic tasks ($F(1, 142) = 19.48$, $MSE = 12.25$, $p<.001$) however after Permission training, p only responses had decreased for both problem types but much more for the arbitrary tasks resulting in little difference between the two problem types ($F(1, 142) = 0.71$, $MSE = 0.44$, $p>.1$). See Figure 5.9 for the graph of the interaction.

	Problem Type	Control group	Permission training
P response	Deontic	0.29	0.07
	arbitrary	0.87	0.18
P/q response	Deontic	0.47	0.43
	arbitrary	2.57	2.86

Table 5.19 Mean p and p/q responses under each condition (score out of 4)

A 2 x 2 (condition x problem type) ANOVA was performed on p/q responses which failed to find an effect of condition ($F(1, 142) = 0.58$, $MSE = 1.12$, $p>.1$) but did find an effect of problem type ($F(1, 142) = 213.02$, $MSE = 369.01$, $p<.001$). In other words there was no difference in p/q responding after training but there was more p/q responding on the arbitrary tasks than deontic. No significant interaction was obtained ($F(1, 142) = 1.15$, $MSE = 2.00$, $p>.1$).

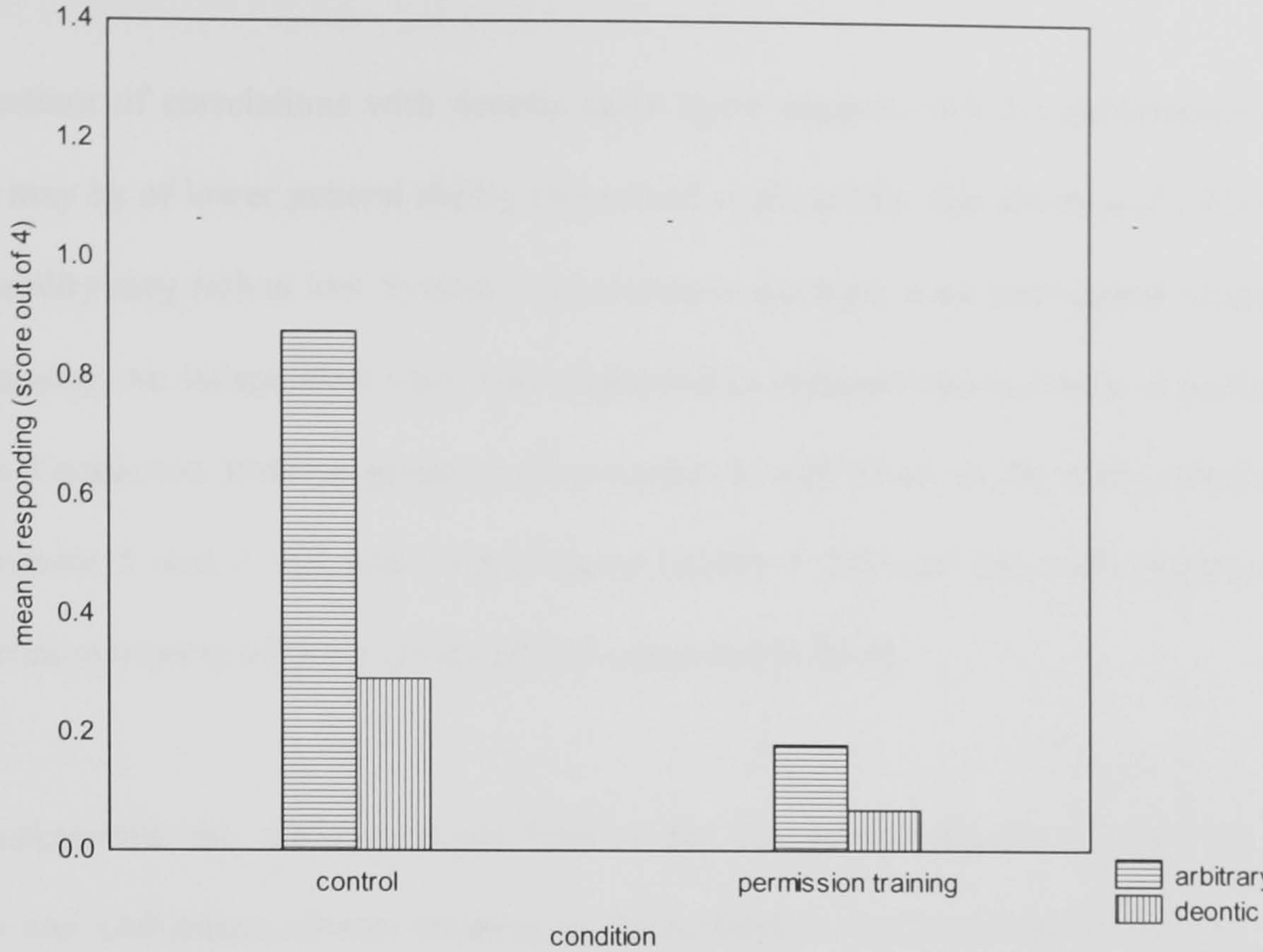

Figure 5.9 Interaction between condition and problem type for p responding

5.11.3 Cognitive ability

Table 5.20 displays the correlations between ability and performance on both types of selection task under each condition. In Experiment 5 there were strong correlations between performance on the deontic and arbitrary problems and ability under Permission training but in Experiment 6 this pattern was absent. In the control condition there was a significant correlation between ability and performance on the deontic problems (.34, p<.05 two tailed) which was present in Experiment 5 prior to Obligation schema training with a lower ability group of participants.

	Control	Permission training
	AH4	
Deontic	.34*	.22
Arbitrary	.21	.12

* = p<.05 **p<.01 (two-tailed)

Table 5.20 Correlations between Ability and performance on the Selection Tasks.

The pattern of correlations with deontic tasks again suggests that the participants in this study may be of lower general ability (Newstead et al., 2004). The absence of correlations with ability may reflect low System 2 involvement amongst most participants in applying the training. An independent t-test was performed to compare ability levels of participants in the Permission training group in Experiment 6 with those in the same condition in Experiment 5 and it did attain significance ($t(100) = 2.25$, $p<.05$) with participants in Experiment 6 being of lower ability (88.75 compared to 96.4).

Consistent with the explanation put forward for the weaker transfer effects found after Rules and Obligation schema training in Experiment 5, we can propose that due to the lower ability levels of the participants in Experiment 6, the Permission training was less effective than in Experiment 5.

5.12 Summary - Experiment 6

With carry over effects and order of presentation effects controlled for, the findings in Experiment 6 support the previous findings that Permission schema training transfers to problems outside the domain of training. However the training effects are weaker than found in Experiment 5. The lower ability levels of the participants in this study and the absence of the correlations between performance on the tasks and ability again lead us to the conclusion that training was not as effective in engaging System 2. These findings will now be discussed as part of the general discussion in this chapter.

5.13 General Discussion - Chapter 5

There were two general predictions made at the beginning of this chapter in relation to performance after abstract training. Evans (1989) has argued that it is doubtful that verbal instruction in relation to underlying logical principles can remove biases as they are implicit and few people get past a relevance judgement on the selection task. In contrast and consistent with Cheng et al.'s and Klaczynski et al.'s (1989; Klaczynski & Laipple, 1993) findings in relation to abstract rules training, it was predicted that a domain-independent rule would be induced from the training which would then be applied to both arbitrary and deontic problems.

In terms of schema-based training there are two contrasting claims in the literature. Consistent with Cheng et al.'s original findings, it was predicted that schema training would transfer to different problem types. However, in accordance with Klaczynski et al.'s dual process account, the training would improve reasoning on problems designed to elicit the schema only.

The pattern of results was not entirely consistent with the previous findings. In Experiments 4 and 5 two training procedures were tested, originally used by Cheng et al. (1986) to investigate transfer effects after training in the logic of the material conditional rule system and after training in a pragmatic reasoning schema. Cheng et al. found that the logic training transferred to selection tasks involving arbitrary and thematic content. However their training involved the use of thematic training examples, which it is proposed, may have facilitated the solution of subsequent thematic problems. They also reported that training in the obligation schema transferred to selection task problems consisting of obligation schema content, but also transferred to two arbitrary problems which they explained may have been interpreted as obligations after training. The two arbitrary problems that may have been interpreted as obligations in Cheng et al.'s study were highly thematic.

In Experiments 4 and 5, after the logical rules training the pattern of results was entirely inconsistent with those predicted by Cheng et al. and Klaczynski et al. who have proposed that training in an abstract rule system would transfer to problems in other domains. In these studies, participants were able to transfer the knowledge to arbitrary problems only. The main effect of training was to reduce p/q responding on the arbitrary tasks. The most obvious explanation for the conflicting results is the change of content in the examples used as part of the training. The transfer that Cheng et al. reported must have at least partly been due to the thematic content of their examples because in both the experiments reported here, this facilitation was absent.

The findings in relation to obligation schema training are consistent with Cheng et al.'s original (1986) study. Transfer occurred to deontic and arbitrary problems though the effects were weaker in Experiment 5. Arguably this finding is entirely inconsistent with dual process theories and Klaczynski et al.'s view that domain-specific training does not transfer outside the domain of training because the information is tied to contexts and experiences and so it cannot be generalised. However it could be argued that the training is not domain-specific any more than the rules training is domain-general. The correlations between intelligence and post-training performance leads us to conclude that explicit training serves to engage System 2 thinking and participants applied the analytic strategies that they have been taught. Individuals were able to decontextualise the key elements of the problem and transfer them to different problems regardless of content.

In Experiments 5 and 6 the training effects using Klaczynski et al.'s (Klaczynski, Gelfand & Reese, 1989; Klaczynski & Laipple, 1993) Abstract and Permission schema training were examined. Again the findings were inconsistent with previous research. Massive facilitation of reasoning responses was found after the examples-based training procedures. The Abstract training transferred to both deontic and arbitrary problems as expected. However the Permission schema training also transferred to both types of selection task, albeit a weaker effect of cross-domain transfer in Experiment 6. In Experiment 5, transfer effects under both training conditions were greater within the same domain as training but large facilitation of correct responding was achieved on the problems from outside the domain also.

There are three key questions to address following the findings reported in this chapter. Firstly why do the findings in relation to Cheng et al.'s rules and obligation schema training conflict with dual process predictions and previous research; secondly why did we find a greater facilitation effect using Klaczynski et al.'s permission schema and abstract training than they did; and thirdly why was Klaczynski et al.'s training so much more effective than Cheng et al.'s? I will address the conflicting findings using Cheng et al.'s training first.

The lack of transfer from the Rules training to deontic problems was not expected. Participants are being taught an abstract rule which has no contextual information which may influence responding or may be mapped onto a specific domain. The general prediction is that participants should be able to map the elements easily onto other problems regardless of content. However both Experiments 4 and 5 found that transfer did not occur. One explanation for this is that the Rules training is difficult and involves reading and understanding quite a complex logical rule system. If we consider the relationships between performance on the tasks after training and cognitive ability into account it is highly likely that only higher ability individuals are able to follow it and then apply it even to problems in the same domain. This is consistent with Sloman's view that individuals of higher ability are more likely to learn the rules (personal communication: 22nd Sept. 2004). Of course it is also possible that the difference in training effects found here compared to Cheng et al.'s original study is due to ability levels. Cheng et al.'s participants may have been higher in ability thus resulting in more positive transfer rates.

Cheng et al. argued that the material conditional was not an intuitive rule system because neither their rules nor examples training alone facilitated reasoning performance. Only when the two procedures were used together was reasoning performance enhanced. The findings reported here support this notion as it would be expected that if the material conditional was intuitive, then after being taught the rule, participants would be able to utilise it on further selection tasks involving both arbitrary and deontic content. The lack of transfer effects in Experiments 4 and 5 casts further doubt on whether the conditional is really material.

In terms of the Obligation training effects in Experiments 4 and 5, transfer occurred to both deontic and arbitrary problems. Consistent with Cheng et al.'s view, participants were able to interpret arbitrary selection tasks as obligation schemas, thus resulting in the correct solution of the task. The difference between Cheng et al.'s training and the training in the experiments reported here is that we included two obligation schema example problems. These may have facilitated solution rates on the arbitrary and deontic test items, or the transfer effects could be due to the provision of the checking procedures as Cheng et al. suggested. Two ways that they suggested that schema training could improve performance were firstly by providing participants with more general mapping rules for interpreting situations in terms of the obligation schema and secondly by providing checking procedures consistent with the material conditional that may be applied to other problem types such as those involving arbitrary content.

Consistent with the Rules training, there were relationships between ability and performance on the arbitrary tasks only after Obligation schema training. Therefore it seems that this training was more effective for higher ability participants when solving arbitrary problems. However, the training facilitated improved reasoning on deontic problems also which could indicate that the schema-based training was more easily transferable to further tasks designed to elicit a pragmatic reasoning schema, regardless of ability levels. This finding is inconsistent with those of Klaczynski and Laipple (1993) who claimed that intelligence is related to the ability to transfer between domain-specific schema problems. Only individuals of higher ability are able to identify the structural similarities between the problems.

In Experiment 5 the Rules and the Obligation schema training were less effective. The explanation put forward for this is that due to the lower ability levels of the participants in Experiment 5 the training was not as effective. The lack of association between performance and ability after both types of training in Experiment 5 and the poorer overall performance indicates that there was less System 2 involvement. Thus, participants in Experiment 4 had a better understanding of the training and could apply it whereas participants in Experiment 5 could not.

Turning now to the massive facilitation effects resulting from Klaczynski et al.'s training procedures, why is Klaczynski et al.'s training procedures more effective in increasing correct solution rates on the selection tasks than Cheng et al.'s? Klaczynski et al.'s training involved presenting participants with four examples followed by in-depth explanation of the correct and incorrect choices. We propose that it is the examples plus feedback themselves that facilitate most of the transfer effects found under all training conditions. In addition, there were only two example problems for each of Cheng et al.'s Rules and Obligation training whereas under Klaczynski's Abstract and Permission training, participants had four problems to acquire the solutions to be used on subsequent selection tasks. All participants had to do in these conditions was recognise that each problem had the same solution regardless of content, understand how the problems were solved and then map the key elements on to the test items. Both the Obligation schema training and the Rules training had complicated explanations of the rule that participants were requested to learn prior to the examples. These explanations are difficult even for higher ability participants to follow and actually may even distract from the goal of the instruction. Klaczynski's training conditions are more effective without these training procedures included.

The effects we found using Klaczynski et al.'s training procedures were inconsistent with the ones they originally reported. The only explanation for this could possibly be that the feedback participants were given during training in Klaczynski et al's studies was presented verbally. Participants in the studies reported here were presented with written explanations of the correct and incorrect card choices, albeit in exactly the same format as Klaczynski et al. used. We propose that the written presentation allowed participants to take in the information and understand it, thus resulting in more transfer from permission to arbitrary problems.

In terms of ability, Klaczynski and Laipple argued that there would be no associations between performance and ability after abstract training as a general problem-solving rule would be acquired to solve subsequent problems. It would be easy to map the rules onto other problems as there would be no content to separate from the key elements in the training. In the Abstract training condition in Experiment 5 associations were found between both deontic and arbitrary tasks and ability after the training, which would indicate that only higher ability participants are correctly learning the rules and applying them.

Klaczynski and Laipple argued that permission training would only improve reasoning on problems from the same domain and individuals that successfully transfer the information would be higher in ability. In contrast with Klaczynski and Laipple, training in a permission schema resulted in improved performance on both deontic and abstract problems. High ability was associated with performance on both types of problems in Experiment 5 which again would indicate that it is higher ability individuals that are able to separate the rule from its context, learn the rule and then apply it to subsequent problems from other domains.

The transfer effects after Permission training found in Experiment 5 were replicated in Experiment 6. Again the participants were lower in ability than the previous group, which would account for the reduced transfer effects. The same explanation can be put forward as for the findings in relation to the reduced transfer effects found after Rules and Obligation schema training in Experiment 5. The Permission schema training was not as effective due to the lower ability group of participants in Experiment 6. The absence of an association between ability and performance after training is consistent with the notion that training is not effectively engaging System 2.

5.14 Conclusion – Chapter 5

The findings from the three experiments reported in this chapter are complex. The simple predictions derived from dual process theories were not supported. For instance it was predicted that no transfer would occur from schema-based training to arbitrary tasks. However, transfer occurred from both Cheng et al.'s and Klaczynski's schema-based training procedures. There are a number of explanations why this might be the case as we have discussed. Further discussion of these finding will be returned to in Chapter 6 which will provide a general discussion of the findings reported in Chapters 3, 4 and 5 of this thesis.

CHAPTER 6

General Discussion

6.1 Introduction

The primary aim of this experimental programme was to investigate the extent to which heuristic and analytic processes affect performance in human thinking, using instruction and training procedures designed to improve reasoning competence. One study examined the performance patterns on a variety of reasoning and decision-making tasks that are documented in the literature for the elicitation of differential responses according to heuristic and analytic processes. The remaining five studies reported have presented evidence to demonstrate that analytic reasoning and decision-making performance can be improved by different training techniques. The aim of Chapter 6 is to discuss the findings of Experiments 1 to 6 and their implications. First we will begin with a summary of the three experimental chapters in relation to their aims and results. Then the implications of the findings will be discussed in terms of the theoretical issues presented in Chapter 1. This will be followed by a consideration of future directions for this research before the final concluding comments.

6.2 Summary of Experimental Findings

In Chapter 3 an individual differences experiment was presented which had three main aims. Firstly, to examine the relationships between statistical and deductive reasoning in order to explore Stanovich's (Stanovich & West, 1998b; Stanovich, 1999) proposal that normative responding on the different problems is related. Secondly, to investigate the relationships between performance on the tasks and individual differences measures of cognitive ability and thinking dispositions. Stanovich (1998a; 1998b) argued that decontextualised reasoning was related to ability whereas Klaczynski, Gordon, and Fauth (1997) proposed that this skill was associated with thinking dispositions. In their view, ability predicted the sophistication of analytic reasoning responses only. Finally, the third aim of the experiment was to develop the law of large numbers and selection task materials for the training experiments reported in Chapters 4 and 5.

All the tasks chosen for Experiment 1 were documented in the literature for their elicitation of differential responding according to analytic and heuristic processes. Both the law of large numbers reasoning and experiment evaluation problems involving manipulations of belief had been shown to elicit two different types of responding cued by logic and belief (Klaczynski, 1997; Klaczynski & Fauth, 1997; Klaczynski, Gordon, & Fauth, 1997;). Performance patterns on both the arbitrary and deontic selection tasks had been interpreted using the two processes, supported by individual differences in cognitive ability and thinking dispositions (Newstead, Handley, Harley, Wright, & Farrelly, 2004; Stanovich & West, 1998a; 1998b). Findings in relation to belief bias in syllogistic reasoning and performance on Stanovich and West's (1997) Argument Evaluation Test had also provided evidence for the existence of the two independent systems.

According to Stanovich and West (1998b; 1999) the ability to decontextualise on a variety of reasoning tasks is linked which indicates that it is a domain-general dispositional trait related to the ability to reason independently of prior beliefs. Participants in their study were presented with a range of reasoning problems including selection tasks, syllogisms, statistical reasoning and Stanovich and West's (1997) Argument Evaluation Test. All the tasks, under a dual process account, are proposed to induce conflict between the two systems of logic and belief. They found that all the tasks correlated with each other which indicated that those participants who respond normatively on one type of task tended to respond normatively on the other.

Experiment 1 replicated and extended on the work of Stanovich and West by introducing the belief-motivated critical reasoning problems as used by Klaczynski, Gordon, and Fauth (1997). Overall, the pattern of correlations indicated that normative responding on the tasks was related. However more associations were predicted than actually obtained. We proposed that the absence of a correlation was not conclusive due to the low number of participants in the study. This was more likely to be a function of statistical power.

It was also predicted that biases elicited on the tasks would be associated with each other. However, no associations between the biases elicited on the different tasks were found. In addition to low power we suggested that it was because biased responses are cued by System 1 processes. These processes are highly contextualised and domain-specific, therefore it would not be expected that bias on different tasks was related. Each task would elicit bias from different domains. One aspect of dual process accounts of the kind Stanovich and Evans and Over proposed is that System 1 maps on to a single cognitive mechanism but of course System 1 effects include such things as matching bias which could be argued as an attentional effect, and belief bias which is a highly contextualised pragmatic effect associated with accessing information in long-term memory. Just because something is labelled as System 1 bias does not mean the same cognitive mechanism produces it. In recognition of this Stanovich has recently referred to System 1 as The Automomous Set of Systems (TASS: Stanovich, 2004). Consequently this lack of correlations between diverse measures of bias is not surprising.

Under a dual process account of reasoning, decontextualised reasoning is an ability associated with System 2's analytical thinking strategies. Stanovich and West (1998b) and Klaczynski, Gordon, and Fauth (1997; Klaczynski & Fauth, 1997) have opposing views on which individual differences factors facilitate this higher level reasoning skill. Stanovich and West argue that cognitive ability is the strongest predictor, in contrast Klaczynski et al.

argued that thinking styles are a better predictor of bias in thinking tasks. Cognitive ability predicts the sophistication of reasoning responses only. We tested these opposing views in Experiment 1 by utilising the same tasks and individual differences measures as both Stanovich and West, and Klaczysnki et al. Namely, the Thinking Dispositions Questionnaire (Stanovich & West, 1998b) and the Rational-Experiential Inventory (Epstein, Pacini, Denes-Raj & Heier, 1996).

Cognitive ability was associated with performance on the experiment evaluation problems and selection task performance, which is consistent with both Stanovich and Klaczynski's findings. Stanovich (1999) proposed that it is the higher ability participants that are able to resist the heuristic (p/q) response on the arbitrary tasks. Klaczynski et al. found that ability was related to performance on experiment evaluation problems. There was also an association between ability and deontic reasoning which is in accord with findings of both Newstead, Handley, Harley, Wright, and Farrelly (2004) and Klaczynski (2001). No further relationships were obtained.

Several associations were found between the thinking styles measures and performance on the tasks, in support of both Stanovich's and Klaczynski's views. For example, the Thinking Disposition Composite score was related to arbitrary selection task responding and objective performance when responding on the Argument Evaluation Test, as would be predicted by Stanovich. It was also found that more open-minded participants were less biased when evaluating strength and persuasiveness of everyday critical reasoning arguments, which is consistent with Klaczynski's (1997; Klaczynski, Gordon, and Fauth, 1997) findings. Again the lack of relationships found could be a function of low statistical power. Several studies have provided evidence for the validity of these thinking styles measures, e.g. Pacini and Epstein (1994). However, no associations have yet been found between performance on reasoning tasks and the experiential subscale of the REI. We propose that you cannot measure implicit System 1 processes using a self-report measure.

In terms of replicating the findings in the literature in relation to analytic and heuristic responding on the tasks, we were quite successful even with a relatively low number of participants. Experiment 1 provided support for the reliability and validity of Stanovich and West's (1997) Argument Evaluation Test, which provides support for their proposal that the AET is a measure of the ability to reason in situations in which prior beliefs may be interfering. The task successfully identified two different groups of participants. One group relied more on objective argument quality for argument evaluation decisions and the other relied more on their prior beliefs. However, no associations were found between objective argument quality and ability as Stanovich and West have reported.

In syllogistic reasoning the typical belief bias effects were found. More valid conclusions were endorsed than invalid and more believable conclusions were endorsed than unbelievable, irrespective of their logical validity (Evans, Barston, & Pollard, 1983; Newstead, Pollard, Evans, & Allen, 1992). The pattern of results in relation to selection task responding was also consistent with previous findings in the literature. Problems consisting of deontic content yielded a higher number of correct solutions than problems involving arbitrary content.

The findings in relation to the belief-motivated critical reasoning problems were not as strong as expected. Participants changed their strategies dependent on whether the conclusion presented was enhancing, threatening or neutral to an individual's occupational goal. More sophisticated reasoning strategies were utilised on problems involving goal-threatening conclusions than goal-enhancing or neutral conclusions but the effects were either very small or not significant. We proposed that the occupational goal manipulation was less effective for our population than for the American population in Klaczynski's studies because the occupational goal that participants had indicated was not a strongly held desire. Therefore participants were not motivated to defend their beliefs.

In Chapter 4, two experiments were reported which were designed to investigate the effects of training in statistical principles on law of large numbers reasoning and bias. This research was inspired by the statistical training studies conducted by Fong, Krantz, and Nisbett (1986) and the work on individual differences in critical reasoning performed by Klaczynski and associates (Klaczynski, 1997; Klaczynski & Gordon, 1996; Klaczynski, Gordon, & Fauth, 1997; Klaczynski & Fauth, 1997).

Fong et al. found that training participants on the law of large numbers rule improved statistical reasoning performance on a variety of everyday reasoning problems. These effects had been replicated using more tightly defined domains although some domain-specificity of the training effects was found with a two-week delay between training and testing (Fong & Nisbett, 1991). In their individual differences research, Klaczynski et al. found that individuals utilise different reasoning strategies dependent on whether the conclusions were congruent or incongruent with their prior beliefs. When the conclusion is belief-congruent, people tend to process the information with little effort and tend to agree with the information. However when the conclusion is belief-incongruent, they tend to utilise more sophisticated reasoning strategies in order to discredit the evidence.

The primary aim of the research presented in Chapter 4 was to examine whether training or instruction procedures would have any impact on belief-based responding. In dual process terms, would training on the explicit system have any effect on the heuristic responses cued by System 1? The results were surprising. It was predicted that at most biased responding would be reduced, but not eliminated by training. Evidence in the belief bias literature indicates that instruction may attenuate but not eliminate belief-based responses. However, the effect of statistical training was to increase the sophistication of statistical reasoning responses on all problem types, whether they were belief-congruent or belief-incongruent. Bias in individuals' responses was eliminated in people's written evaluations of the arguments.

However, this effect was not reflected in the rating scales. These scales required participants to indicate on two nine-point scales how strong they thought the conclusion was based on the evidence presented, and how convinced they were by the argument. After the training, even though participants were able to utilise statistical principles in their written justifications of why the arguments were strong or weak, they still rated them in accordance with their beliefs. That is when the argument was belief-consistent, it was rated as strong and more convincing. When the argument was belief-inconsistent, it was rated as weak and unconvincing. This demonstrates a dissociation between analytic and heuristic responding and we will consider this further in a later section.

In Experiment 3 the findings from Experiment 2 were replicated. An instruction technique was included, based on the ones used in the syllogistic reasoning literature, to examine whether participants could disengage from their beliefs when asked to. Participants were instructed to base their evaluations on the content of the arguments, not what they believed to be true. This resulted in no increase in analytic responding at all. It appears that people require more than motivational instructions to separate prior beliefs from the logic of the problem. As discussed previously, being presented with one piece of evidence that is inconsistent with your belief will have a minimal impact in terms of changing that belief according to Bayes theorem. However there were indications that System 2 was being engaged in some way. Participants rated the evidence as stronger and more convincing after instruction which may indicate that they are interpreting the instructions as a reason not to evaluate the evidence strength as low. Perhaps the instruction to assume the information was true caused participants to have more faith in the veracity of the evidence and hence conclude that the arguments must be stronger. Again this will be discussed further later in this chapter.

In Chapter 5, a series of three experiments were reported which investigated the effects of training on the selection task. Experiment 4 was designed to extend on the work of Cheng, Holyoak, Nisbett, and Oliver (1986). They proposed that training on the logic of the material conditional improved reasoning on selection task problems involving arbitrary and thematic content, but only when examples were given as part of the training. Cheng et al. also found that training in the obligation schema (Cheng & Holyoak, 1985) facilitated reasoning performance on subsequent selection tasks involving both arbitrary and obligation schema content.

It is important to know whether there is transfer or not from the different training procedures because according to the dual process theorist Klaczynski, abstract rule training would transfer to problems involving thematic content. However domain-specific training such as obligation schema training would not transfer to problems outside of that domain.

There were confounds in their research that we needed to address first in order to be able to make claims about the effects of explicit training on the analytic and heuristic processes that cue responses on the selection tasks. Firstly, the examples that were used as part of their rules training were thematic which may have facilitated the transfer to further thematic problems. These were changed to tasks consisting of arbitrary content. Secondly, Cheng et al. found that their obligation schema training improved reasoning on two arbitrary problems which they suggested may have been interpreted as obligations after the training. The two test items that they identified were also thematic which may have facilitated their solution. Participants in Experiment 4 were presented with either abstract rules plus examples training or obligation schema training. They were all given the same four deontic and four arbitrary selection task problems for the test. We found that the abstract rules training transferred only to problems consisting of arbitrary content, and the obligation schema training transferred to both arbitrary and deontic tasks, in both Experiments 4 and 5.

These findings are inconsistent with Cheng et al. and with a simple interpretation of dual process theories. Under a dual process account, training or instruction in a System 2 domain-independent rule, such as the rules of logic, should transfer to problems from other domains. In contrast, domain-specific training i.e. obligation schema training, should only facilitate reasoning on tasks from the same domain. Due to the conflicting findings, Cheng et al.'s training procedures were directly compared to Klaczynski, Gelfand, and Reese's (1989; Klaczynski & Laipple, 1993) training procedures, whose findings had previously been interpreted under dual process accounts. However, we found conflicting results using Klaczynski et al.'s training also. Consistent with dual processes, large effects were found with their abstract training on both arbitrary and deontic problems. However, we found their permission schema training transferred to both types of problem, arbitrary and deontic, also.

The complexity of the different training procedures was the explanation put forward for the large differences between Cheng et al.'s and Klaczynski's transfer effects. Cheng et al.'s abstract rules and obligation schema training commenced with detailed descriptions of the rules they were to learn. In contrast, Klaczynski's training involved the presentation of four individual selection task problems followed by explanations of the correct and incorrect choices to be made. All participants had to do was repeat the strategies on further selection tasks. There were some domain-specificity effects of training i.e. participants in the abstract training condition performed better on the arbitrary test items and participants in the permission group performed better on the deontic items.

Findings in relation to individual differences, specifically cognitive ability were particularly interesting in this series of studies. The difference in ability levels between the groups of participants appeared to reflect the effectiveness of the training procedures. In Experiment 5 the rules and obligation training effects were not so large with a lower ability group compared to Experiment 4, and in Experiment 6 the permission training effects were small compared to Experiment 5, again with a lower ability group of participants. In addition, associations between performance on the selection tasks after training and ability were absent with lower ability participants. It was suggested one way of interpreting correlations between ability and performance is as an indicator of System 2 involvement in a task. The absence of correlations and the weaker training effects, led us to the conclusion that System 2 was not as engaged by the training for the lower ability groups. Training was more effective for higher ability participants as they had more cognitive resources to draw upon. Further discussion of these findings will be presented in the next section of this chapter.

6.3 Theoretical Implications

In this section we will consider the theoretical implications of the findings from this thesis in relation to three main themes. Firstly, what the findings say about the effectiveness of training, secondly the implications of the findings for dual process theories, and thirdly the implications of the findings for other theories of reasoning.

In Chapter 1 of this thesis, several questions were posed in relation to the effectiveness of different training techniques. Firstly would it be possible to improve people's critical thinking skills, and if so which training would be the most effective in doing so? Secondly, is it possible to debias people's thinking, or reduce their biased responding? Finally, if individuals are taught a skill or a strategy on one type of problem, can this then be transferred to other types of problems consisting of different content than the one the individual was taught in? These questions have all been addressed in this programme of research.

6.4 The Effectiveness of Training

6.4.1 Can critical thinking be improved?

Earlier training research provided evidence to show that, in accordance with Plato's formalist tradition, quality of reasoning can be improved by the teaching of general inferential rules (Lehman, Lempert, & Nisbett, 1988; Lehman & Nisbett, 1990). These rules can then be utilised on other types of problems outside the domain of training. However, a problem with this research was that it was too general. Participants were recruited from different courses and then their ability measured on different types of reasoning: statistical, methodological and conditional. Lehman et al.'s conclusions were based on the rules that supposedly were taught implicitly by the courses involved. But there is no way of controlling for the precise factors which facilitated this improvement, or being sure of the rules that are being taught. The findings in the current research extend on this by being far more specific about the rules and strategies being taught. In this way we can also be far more specific about our conclusions.

All the training procedures reported here were found to improve critical reasoning skills in the short term. Training on the law of large numbers increased analytic reasoning performance on a variety of everyday reasoning problems, and the schema-based training facilitated reasoning on the selection task. Logic training also increased correct responding on selection task problems although the effects were not so strong as the other training procedures. The results indicate that given the appropriate rules and strategies for the task, reasoning performance on that task will be enhanced.

There are indications though from the selection task studies reported in Chapter 5 that ability is a key factor when looking at receptiveness to different training techniques. The findings demonstrate that participants of higher ability appear to be more able to understand and apply the rules and strategies being taught. Training was not as effective for participants of lower ability. Whether this would be the same for other types of reasoning and training procedures it is difficult to say. Further investigation is required before generalisations of the findings reported here could be made.

6.4.2 Can training debias?

Can we conclude that training is effective in removing bias? Apart from in this programme of research, the only research performed that has demonstrated the positive effects of instruction on bias is the syllogistic reasoning literature (Evans, Newstead, Allen, & Pollard, 1994; Newstead, Pollard, Evans, & Allen, 1992), the findings of which have been the source of many of the predictions made for the experiments reported here. It was found that belief bias could be reduced but not eliminated. The findings reported here in relation to the law of large numbers were far more dramatic. Bias was eliminated on the written evaluations in Experiments 2 and 3. Klaczynski has suggested that the training may not even transfer to problems involving manipulations of belief (personal communication, Nov. 2004). Sloman suggested that the training would not transfer to problems very different from the training ones (Sloman, personal communication, September 2004). The findings here demonstrate that given the training on the law of large numbers, people are able to utilise the rule effectively, therefore overriding their belief-based response.

However, the training did not decrease bias in ratings of the arguments, therefore demonstrating the pervasiveness of beliefs. But why was the bias still present on the rating scales? Participants had received the statistical training and were able to utilise the LLN rule as evidenced in their written justifications. One explanation is the amount of effort the participants were required to use on the different response formats. When responding on a rating scale, participants may provide a quick response. In contrast a written evaluation requires more effort and elaboration of why the arguments are strong/weak or convincing/unconvincing. The two types of response can be mapped onto the different responses cued by the two systems in terms of dual process accounts. When no effort is required the influence of System 1 belief-based reasoning is observed in the ratings. In this case, if a person has a high degree of belief in a claim, one piece of evidence that contradicts this belief will not have much impact. Thus bias is still present in the ratings. However, when a written justification is requested, participants expend more effort and utilise analytic System 2 strategies leading to the inhibition of System 1 responses. Thus a dissociation between the responses cued by the two systems is observed.

The selection task studies support this conjecture. The heuristic response on arbitrary selection tasks is not due to belief as in the LLN problems. However Evans (1989) argued that few people get past a relevance judgement on this task and the cards are selected pre-attentionally. Training served to engage analytic reasoning and provided participants with the strategies to utilise on the tasks, thus resulting in a decrease of heuristic responses. Overall, training is successful in reducing biased responding on both LLN problems and the selection tasks, specifically when the strategies for the correct response are provided.

Instruction to dissociate from beliefs had no effect on biased responding on law of large numbers problems. Again this demonstrates the pervasive influence of belief on reasoning. Klaczynski and Gordon (1996) also found no effect of motivational instruction on bias.

This illustrates yet again that people need to be provided with the strategy or rule so that they can deal with the problem analytically. Sloman (personal communication, September 2004) proposed that people need to be cued to use a rule and that training will be effective in reducing bias to the extent that the cue is effective. The findings from both the statistical and selection task studies demonstrate that critical reasoning can be improved and bias reduced or eliminated provided the appropriate rules and strategies are taught. Participants are responding to the instructions, albeit in a different way to what was expected, as they rated the evidence as stronger and more persuasive. This is possibly because they are assuming that the information is true (as told to in the instructional set) and thus more reliable, which leads participants to provide a more positive evaluation of the evidence.

To summarize, the training is effective in reducing bias and in the case of LLN reasoning, the effect is quite dramatic. It appears that providing people with the strategies enables them to utilise them on different problems, which subsequently impacts on the biased responses. The extent to which the bias is reduced is probably due to the complexity of the strategy (or rule) that they have to learn. It would be very interesting to provide participants in a study of belief bias, training on how to successfully follow the logic of syllogistic reasoning. We would predict from the findings reported here, that given the strategies participants would be able to transfer the knowledge on to other problems consisting of belief-based content. Thus resulting in an elimination of bias, rather than reduction.

6.4.3 Can the skills be transferred?

All the training procedures employed in this thesis resulted in an improvement in reasoning performance. However they varied in the extent to which the skills taught could be transferred to other problems. Training in statistical principles improved reasoning across domains, but transfer after selection task training was highly dependent on the type of instruction received. Transfer across domains occurred after Klaczynski's abstract training, the permission schema training, and Cheng et al.'s obligation schema training. In contrast, participants were able to utilise Cheng et al.'s logical rules training on further arbitrary problems only. This was the only training procedure that people were not able to transfer to a different type of problem.

What was it about the rules training that was different from the rest? All the training procedures involved examples as part of the training. One explanation we propose is that use of examples facilitated reasoning on subsequent problems (for LLN and selection tasks). In accordance with Klaczynski, Gelfand, and Reese (1989) people were able to recognise the key elements of the problems, and transfer them to other problems. However, rules training consisted of two abstract training examples also. According to all the predictions it should have been possible to map the elements on to subsequent tasks easily, whatever the problem content. However it is possible that the explanation of the rules of the material conditional was just too difficult for participants to understand. Cheng et al. originally suggested that the material conditional is not an intuitive rule based on their findings that participants' reasoning performance was not improved after rules training alone. If the rules training does not map on to an intuitive rule then it would be like teaching participants the laws of physics or a new artificial language. Participants may have been able to utilise some of the elements on further arbitrary tasks but not generalise them to thematic ones. Supporting evidence for this view can be found in recent work on conditionals that claims the material conditional is not the way that people naturally interpret conditional sentences (Evans, Handley, & Over, 2003).

Klaczynski's abstract training resulted in large effects of training on both abstract and deontic problems yet it only differed from Cheng et al.'s in that it consisted of four examples rather than two, and did not involve the logical rules explanation. It appears that people are much more able to understand examples and practise the procedures, therefore learning by experience, when the rule or strategy is not intuitive.

It could be argued that Fong et al.'s LLN training involved the same procedural format yet resulted in cross-domain transfer. However the LLN rules training is in contrast much simpler to follow. In addition it is proposed that the law of large numbers is an intuitive rule. The fact that reasoning performance is improved by rules training alone (Fong et al. 1986), and people are able to utilise the rule without training if the context of the problem cues it, suggests that people already have a rough version of the rule intuitively. In contrast to the rules training in the selection task studies, the LLN rules explanation is context-bound and involves a demonstration which clarifies the rule being taught for the participants. They are not just learning an abstract rule. All participants have to remember is "the larger the sample, the better" and then they can use this on any problem. The logical rules training is a great deal more complicated which is probably why Fong et al. and Cheng et al.'s original training studies resulted in such conflicting findings. Statistical reasoning in Fong et al.'s study was found to be improved after either the LLN rules or the examples training, whereas selection task reasoning was only improved after the two techniques together in Cheng et al.'s.

Obligation and permission schema training were both successful in cross-domain transfer even though they were not expected to be. Klaczynski et al. argued that transfer would only be unidirectional, from abstract to deontic, not the other way round. Again, the only conclusion can be that the examples facilitated the transfer. With retrospect it would have been a lot more informative if examples had not been included as part of the obligation schema training in Experiments 3 and 4. Then it would have been possible to examine whether Cheng et al.'s original training procedure did result in the transfer to arbitrary problems. By including them we can only suggest that the examples facilitated this transfer due to the findings in relation to Klaczynski's permission schema training.

Even though most of the training procedures resulted in positive transfer to other problem types, there were some domain-specificity effects in the selection task studies, i.e. abstract transferred more to further abstract tasks, deontic more to deontic. According to Klaczynski et al., domain-specific training coincides with the domain's pre-existing rules. Thus transfer occurs to other domain-specific problems but inhibits transfer to arbitrary problems. They found no transfer from permission schema training to arbitrary problems at all. One explanation for the results reported here could be that participants are acquiring the rule in a specific context, therefore when giving the problem in a different context, it is a little harder to decontextualise the key elements and transfer them. In summary, training in reasoning is effective. If people are provided with the strategies they can successfully utilise them on further problems, even in different contexts.

6.5 Implications for Dual Process Theories

Dual process theory has been used as a theoretical framework throughout this thesis. Many of the findings here appear inconsistent with dual process theories. Generally, it was predicted that training would impact on System 2 processes, not System 1. More specifically, in accordance with the belief bias findings, bias may be reduced but not eliminated (Stanovich, personal communication, October 2004; Evans et al. 1994).

The LLN training studies demonstrate how the effects of bias can be eliminated when participants are asked for written analysis of the arguments. Participants were able to utilise the LLN strategies competently in their justifications of why the arguments were strong or weak, and convincing or unconvincing. However simply being asked to indicate their thoughts on a scale led to responses that were influenced by beliefs. When the evidence was consistent with beliefs, participants rated it as stronger and more convincing, but when it was inconsistent with beliefs participants rated the evidence as weaker and less convincing. Thus people appeared to consciously inhibit beliefs in their written justifications. However the ratings reflect the highly pervasive nature of System 1 responses. Importantly for dual process theories, these findings demonstrate a dissociation between responses cued by the two systems, where System 2 engagement appears greatest when participants are asked to externalise their reasoning.

The selection task training also resulted in a decrease of heuristic responding. Even the domain-specific schema training increased analytic responding. According to Klaczynski's dual process theory, one might expect that training in a domain-specific schema would not transfer to problems from other domains (Klaczynski, Gelfand, & Reese, 1989; Klaczynski & Laipple, 1993). However the schema training did transfer to other domains. This is not all bad news for dual process theorists. According to dual process theories training serves to engage System 2 thinking. Whether the training is abstract or tied to a specific domain, people are able to utilise the rule or strategy on other problems. It may be more difficult to transfer across domains judging by the domain-specificity effects noted from the training, but people are able to do so resulting in a reduction of System 1 responses. As Evans and Over (1996) proposed, training increased rationality 2. In doing so, System 1 responses were therefore inhibited or overridden.

The findings in relation to ability demonstrate how System 2 processes mediate the effectiveness of training. According to Stanovich (1999; Stanovich & West, 1998b) higher ability individuals are able to decontextualise from the context of a problem and respond normatively. We have found some evidence to support this in Experiment 1. Also, the selection task studies reported in Chapter 5, indicate that higher ability participants are more receptive to the training. Sloman (personal comm. September 2004) proposed that higher ability individuals are more likely to learn the rule whereas Klaczynski (personal comm. November 2004) suggested that higher ability individuals would acquire the rule more rapidly. We propose that training impacts on System 2, and any associations found after instructional training between performance and ability are an index of System 2 engagement. Those individuals with more cognitive resources are more able to draw on those resources and use them to apply the strategies on further problems. Thus the training is more effective for those individuals.

The findings in this thesis in relation to thinking styles are difficult to interpret. Stanovich suggests that thinking dispositions and ability determine how well someone decontextualises. In contrast Klaczynski, Gordon, and Fauth argue that disposition is the key factor in predicting this skill. No firm conclusions can be made from the findings here. There were no clear interpretable relationships between either of the thinking dispositions measures and reasoning performance or bias. Further research is required before either position can be supported.

One thing that is apparent in relation to the simple predictions made from Klaczynski's dual process accounts is that they are not sufficient to account for the findings reported in this thesis. Clearer predictions need to be developed. The question of how training interacts with System 1 and System 2 processes is a great deal more complicated than what has been proposed so far. The statistical training studies demonstrate that different response formats change the types of influences on responses. Written explanations elicited analytic System 2 responses, whereas rating scales allow belief-based System 1 influences on responses. The selection task studies resulted in completely inconsistent findings to what was expected. Logical rules training did not transfer whereas training in a pragmatic reasoning schema did. The one consolation for dual process theories is that System 2 is sensitive to instruction techniques. There is still a great deal of work to be done in this area.

6.6 Implications for other Theories of Reasoning

What implications do the findings reported here have for other theories of reasoning? The findings in relation to the selection task are consistent with domain-specific theories such as pragmatic reasoning schemas (Cheng & Holyoak, 1985) or social contracts (Cosmides, 1989). People perform better overall (in the absence of training) when the problems are contextualised. However it would be expected that domain-specific training would not transfer to other domains. The training would improve reasoning on problems that could be interpreted within the same domain only. The findings reported in Chapter 5 demonstrate that schema-based training does transfer outside of the training domain. There was some domain-specificity of training noted which indicates that people find cross-domain transfer a little more difficult, but the improvement was still considerable. This has implications of how people acquire domain-specific knowledge in the first place and whether these domains are so tightly defined.

The finding that training in the rule of the material conditional only impacts on arbitrary reasoning problems may have implications for mental models theories. Mental models has been the most popular theory of how people represent conditionals for a long time (Johnson-Laird & Byrne, 2002). Findings discussed in terms of mental models are based on the material conditional as the logical rule system that reasoning behaviour is compared to. Cheng et al. proposed that the material conditional was not a naturally occurring rule system because people were not able to use it after the rules training. If it was an intuitive rule then participants should be able to use it after an explanation of it as with the LLN rule. We find support for this here, as participants were not able to acquire and transfer the abstract rule to other problem types. As suggested earlier in this discussion, participants found the rule difficult to learn as to them it was a completely novel rule, like attempting to learn an artificial language. We propose that from the findings reported here and in addition to Cheng et al., the material conditional does not represent the logic of the conditional in everyday language.

In Chapter 1, as part of the introduction to this thesis we discussed how reasoning research for a long time has been comparing performance against the wrong normative systems of logic. Mental models theory may still be using the wrong logical system. Current research is now looking at alternative logical rule systems which may be more representative of how people reason about conditionals (Evans & Over, 2004; Evans, Over, & Handley, in press).

6.7 Future Directions and Conclusions

The training studies reported here have been very successful in increasing analytic reasoning performance and reducing biased responding. We have considered our findings in the context of a dual process framework, namely the impact of instruction techniques on Systems 1 and 2. Training has successfully increased analytic processing by providing the strategies and cues necessary to engage it. In turn this has led to a decrease in biased responding.

Ability has been found to have a role in mediating the effectiveness of training. Training was more effective for higher ability participants. We propose that the associations found between ability and performance on the tasks after training is an index of System 2 engagement. Those of higher ability are more able to understand and apply the principles that they have been taught. Thus training is more effective for these individuals. However we still do not know anything about the actual processing which takes place within the two systems, either the analytic processing or the intuitive.

Further training and instruction studies are required to investigate the different procedures that facilitate transfer on different types of reasoning problems. The findings from this programme of research indicate that by providing people with the right strategies, reasoning performance can be improved which leads to a reduction in bias. We need to examine whether these results can be generalised to other reasoning and decision making tasks, and if so, what type of training and how long will the effects last.

In terms of dual process accounts it is necessary to understand how different manipulations impact on the two systems, and how the systems interact with each other. By conducting further training studies using different tasks and materials we may gain more insight into the specific processes utilised. One way of investigating the two reasoning systems further using instructional techniques would be an extension of the LLN training studies presented here. By ascertaining participants' beliefs in the conclusions prior to training, and employing individual differences measures it would be possible to investigate the extent to which ability can facilitate decontextualised reasoning and training effectiveness, dependent on the extent to which the beliefs are strongly held. This would provide more information about how the two systems interact with each other in everyday critical reasoning.

There has been a complete shift in the last 30 years in how reasoning and decision-making performance is studied. Research using the deduction paradigm as discussed at the beginning of this thesis, is being replaced by studies which consider the contextual factors which may facilitate or inhibit responding, complemented by research which has an emphasis on individual differences. Errors and biases in judgement are viewed as informative in addition to the normative responses.

The important finding from this programme of research is that critical reasoning can be improved, and the extent of that improvement is due to the specific rules and strategies being taught, relevant to the type of reasoning or decision making task. Dual processes are becoming increasingly popular as a way of conceptualising experimental findings concerning the influence of instruction on training. However this thesis shows that more work is needed before they can be used to make clear and reliable predictions. It is necessary to go beyond a simple alignment of tasks to different processing mechanisms and a simple interpretation of the way in which dual processes may guide reasoning with different types of materials. As this thesis shows, predictions derived from dual process theories concerning the influence of training on thinking are often not borne out by the data.

References

Ackerman, P. L. & Heggestad, E. D. (1997). Intelligence, Personality, and Interests: Evidence for Overlapping Traits. *Psychological Bulletin, 121,* 219-245.

Almor, A. & Sloman, S. A. (1996). Is Deontic Reasoning Special. *Psychological Review, 103,* 374-380.

Ansburg, P. I. & Shields, L. (2003). Training overcomes reasoning schema effects and promotes transfer. *The Psychological Record, 53,* 242.

Bassok, M. & Holyoak, K. J. (1989). Interdomain Transfer Between Isomorphic Topics in Algebra and Physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 153-166.

Bowers, T. L. & Pantle, M. L. (1998). Shipley Institute for Living Scale and the Kaufman Brief Intelligence Test as Screening Instruments for Intelligence. *Assessment, 5,* 187-195.

Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching Thinking and Problems Solving. *American Psychologist, 41,* 1078-1089.

Cacioppo, J. T. & Petty, R. E. (1982). The Need for Cognition. *Journal of Personality and Social Psychology, 42,* 116-131.

Chater, N. & Oaksford, M. (2001). Human rationality and the psychology of reasoning: Where do we go from here? *British Journal of Psychology, 92,* 193-216.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology. 17(4),* 391-416.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus Syntactic Approaches to Training Deductive Reasoning. *Cognitive Psychology, 18,* 293-328.

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

Denes-Raj, V. & Epstein, S. (1994). Conflict Between Intuitive and Rational Processing: When People Behave Against Their Better Judgment. *Journal of Personality and Social Psychology, 66,* 819-829.

Derry, S., Levin, J. R., & Schauble, L. (1995). Stimulating Statistical Thinking Through Situated Simulations. *Teaching of Psychology, 22,* 51-57.

Donovan, S. & Epstein, S. (1997). The Difficulty of the Linda Confunction Problem Can Be Attributed to Its Simultaneous Concrete and Unnatural Representation, and Not to Conversational Implicature. *Journal of Experimental Social Psychology, 33,* 1-20.

Dunbar, G. L. (2000). Traces of reasoning with pragmatic schemas. *Thinking and Reasoning, 6,* 173-181.

Epstein, S. (1994). An integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709-724.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual Differences in Intuitive-Experiential and Analytical-Rational Thinking Styles. *Journal of Personality and Social Psychology, 71,* 390-405.

Evans, J. St. B. T. (1972). On the problems of interpreting reasoning data: Logical and psychological approaches. *Cognition, 1,* 373-384.

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*. Brighton: Erlbaum.

Evans, J. St. B. T. (1993). Bias and rationality. In K.I. Manktelow & D.E. Over (Eds), *Rationality*. London: Routledge.

Evans, J. St. B. T. (2002). Logic and Human Reasoning: An Assessment of the Deduction Paradigm. *Psychological Bulletin, 128,* 978-996.

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. (2004). History of the dual process theory of reasoning. In K.I.Manktelow & M. C. Chung (Eds.), *Psychology of Reasoning: Theoretical and Historical Perspectives* (Hove, UK: Psychology Press.

Evans, J.St. BT & Lynch, J. (1973). Matching bias in the selection task. *British Journal of Psychology, 64*, 391-97.

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

Evans, J. St. B. T., Over, D. E., & Manktelow, K. I. (1993). Reasoning, decision making and rationality. *Cognition, 49,* 165-187.

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

Evans, J. St. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by Instruction: The Case of Belief Bias. *European Journal of Cognitive Psychology, 6,* 263-285.

Evans, J. St. B. T., Barson, J. L. & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition, 11,* 295-306.

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

Evans, J. St. B. T., Handley, S. H. & Over, D.E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 321-355.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The Effects of Statistical Training on Thinking about Everyday Problems. *Cognitive Psychology, 18,* 253-292.

Fong, G. T. & Nisbett, R. E. (1991). Immediate and Delayed Transfer of Training Effects in Statistical Reasoning. *Journal of Experimental Psychology: General, 120,* 34-45.

Galotti, K. M. (1989). Approaches to Studying Formal and Everyday Reasoning. *Psychological Bulletin, 105,* 331-351.

Galotti, K. M. (1995). Teaching Critical Thinking: Reasoning About Reasoning: A Course Project. *Teaching of Psychology, 22,* 66-68.

Gick, M. L. & Holyoak, K. J. (1983). Schema Induction and Analogical Transfer. *Cognitive Psychology, 15,* 1-38.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and Biases.* New York: Cambridge University Press.

Girotto, V., Kemmelmeier, M., Sperber, D., & Henst, J. B. v. d. (2001). Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition, 81,* B69-B76.

Goel, V. & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87,* B11-B22.

Griggs, RA, & Cox, JR (1982). The elusive thematics material effect in Wason's selection task. *British Journal of Psychology, 73*, 407-420.

Griggs, R. A. & Cox, J. R. (1993). Permission Schemas and the Selection Task. *The Quarterly Journal of Experimental Psychology, 46A,* 637-651.

Handley, S. J., Newstead, S. E., & Wright, H. L. (2000). Rational and Experiential Thinking: A study of the REI. In R.J.Riding & S. G. Rayner (Eds.), *International Perspectives on Individual Differences. Volume 1. Cognitive Styles* (ed., pp. 97-113). Stamford, Connecticut: Ablex Publishing Corporation.

Hanley, G. L. (1995). Teaching Critical Thinking: Focusing on Metacognitive Skills and Problem Solving. *Teaching of Psychology, 22,* 68-72.

Henle, M. (1962). On the relation between logic and thinking, *Psychological Review, 69*, 366-378.

Hirt, E. R. & Markman, K. D. (1995). Multiple Explanation: A Consider-an-Alternative Strategy for Debiasing Judgments. *Journal of Personality and Social Psychology, 69,* 1069-1086.

Holland, J.H., Holyoak, K.J., Nisbett, R.E. & Thagard, P.R. (1986). *Induction: Processes on inference, learning and discovery.* Cambridge, Mass.: M.I.T. Press.

Holyoak, K. J. & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15,* 332-340.

Houdé, O. & Moutier, S. (1996). Deductive reasoning and experimental inhibition training: the case of the matching bias. *Current Psychology of Cognition, 15,* 409-434.

Houdé, O. & Moutier, S. (1999). Deductive reasoning and experimental inhibition training: The case of the matching bias. New data and reply to Girotto. *Current Psychology of Cognition, 18,* 75-85.

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B. et al. (2000). Shifting from the Perceptual Brain to the Logical Brain: The Neural Impact of Cognitive Inhibition Training. *Journal of Cognitive Neuroscience, 12,* 721-728.

Inhelder, B. and Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence.* Basic Books, New York, NY

Jakoubek, J. (1995). Developing Critical-Thinking Skills in Psychology Content Courses. *Teaching of Psychology, 22,* 57-59.

Jepson, C., Krantz, D., & Nisbett, R. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences, 6,* 494-501.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition, 50,* 189-209.

Kahneman, D. & Frederick, S. (2002). Representativeness revisisted: attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (eds), *Heuristics and biases: the psychology of intuitive judgement.* Cambridge: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky,A. (1982). *Judgment Under Uncertainty: Heurisics and biases.* Cambridge, U.K.: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Kardash, C. M. & Scholes, R. J. (1996). Effects of Preexisting Beliefs, Epistemological Beliefs, and Need for Cognition on Interpretation of Controversial Issues. *Journal of Educational Psychology, 88,* 260-271.

Keren, G. & Lewis, C. (2000). Even Bernoulli Might Have Been Wrong: A Comment on Intuitions about Sample Size. *Journal of Behavioural Decision Making, 13,* 125-132.

Klaczynski, P. A., Gelfand, H., & Reese, H. W. (1989). Transfer of conditional reasoning: Effects of explanations and initial problem types. *Memory & Cognition, 17,* 208-220.

Klaczynski, P. A. & Laipple, J. S. (1993). Role of Content Domain, Logic Training, and IQ in Rule Acquisition and Transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 653-672.

Klaczynski, P. A., Gordon, D. H., & Fauth, J. (1997). Goal-Oriented Critical Reasoning and Individual Differences in Critical Reasoning Biases. *Journal of Educational Psychology, 89,* 470-485.

Klaczynski, P. A. (1993). Reasoning schema effects on adolescent rule acquisition and transfer. *Journal of Educational Psychology, 85*, 679-692.

Klaczynski, P. A. (1997). Bias in Adosescents' Everyday Reasoning and Its Relationship With Intellectual Ability, Personal Theories, and Self-Serving Motivation. *Developmental Psychology, 33,* 273-283.

Klaczynski, P. A. & Fauth, J. (1997). Developmental Differences in Memory-Based Intrusions and Self-Serving Statistical Reasoning Biases. *Merrill-Palmer Quarterly, 43,* 539-566.

Klaczynski, P. A., Fauth, J. M., & Swanger, A. (1998). Adolescent Identity: Rational vs. Experiential Processing, Formal Operations, and Critical Thinking Beliefs. *Journal of Youth and Adolescence, 27,* 185-207.

Klaczynski, P. A. & Robinson, B. (2000). Personal Theories, Intellectual Ability, and Epistemological Beliefs: Adult Age Differences in Everyday Reasoning Biases. *Psychology and Aging, 15,* 400-416.

Klaczynski, P. A. (2001). The influence of analytic and heuristic processing on adolescent reasoning and decision making. *Child Development, 72*, 844-861.

Klaczynski, P. A. (2004). A Dual-Process Model of Adolescent Development: Implications for Decision Making, Reasoning, and Identity. *Advances in ChildDevelopment and Behavior, 31*.

Klaczynski, P. A., Schuneman, M. J., & Daniel, D. B. (2004). Theories of Conditional Reasoning: A Developmental Examination of Competing Hypotheses. *don't know*.

Klauer, K. C., Meiser, T., & Naumer, B. (2000). Training propositional reasoning. *The Quarterly Journal of Experimental Psychology, 53A*, 868-895.

Klauer, K. J. (1996). Teaching Inductive Reasoning: Some Theory and Three Experimental Studies. *Psycinfo, 6*, 37-57.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107-118.

Kosonen, P. & Winne, P. H. (1995). Effects of Teaching Statistical Laws on Reasoning About Everyday Problems. *Journal of Educational Psychology, 87*, 33-46.

Kuhn, D. (1993). Connecting Scientific and Informal Reasoning. *Merrill-Palmer Quarterly, 39*, 74-103.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin, 108*, 480-498.

Lawson, T. J., Schwiers, M., Doellman, M., Grady, G., & Kelnhofer, R. (2003). Enhancing Students' Ability to Use Statistical Reasoning With Everyday Problems. *Teaching of Psychology, 30*, 107-110.

Lee, Y. & Vakock, D. A. (1996). Transfer and retention of implicit and explicit learning. *British Journal of Psychology, 87*, 637-651.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The Effects of Graduate Training on Reasoning: Formal Discipline and Thinking About Everyday-Life Events. *American Psychologist, 43*, 431-442.

Lehman, D. R. & Nisbett, R. E. (1990). A Longitudinal Study of the Effects of Undergraduate Training on Reasoning. *Developmental Psychology, 26,* 952-960.

Levinson, S.C. (1983).*Pragmatics.* Cambridge, England: Cambridge University

Liberman, N. & Yechiel, K. (1996). Hypothesis testing in Wason's selection task: social exchange cheating detection or task understanding. *Cognition, 58,* 127-156.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098-2109.

Manktelow, K., & Evans, J.St.BT (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.

Manktelow, K. I. & Over, D. E. (1993). *Rationality: Psychological and philosophical perspectives.* London: Routledge and Kegan Paul.

Manktelow, K. I., Sutherland, E. J., & Over, D. E. (1995). Probabilistic Factors in Deontic Reasoning. *Thinking and Reasoning, 1,* 201-219.

McGuinness, C. & Nisbet, J. (1991). Teaching Thinking In Europe. *British Journal of Educational Psychology, 61,* 174-186.

McGuinness, C. (1993). Teaching Thinking: new signs for theories of cognition. *Educational Psychology, 13,* 305-316.

McMillan, J. H. (1987). Enhancing College Students' Critical Thinking. *Research in Higher Education, 26,* 3-29..

Moutier, S. (2000). Deductive reasoning and experimental matching-bias inhibition training in school. *Current Psychology of Cognition, 19,* 429-452.

Newstead, S. E., Pollard, P., Evans, J. St. B. T., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition, 45,* 257-284.

Newstead, S. E. & Evans, J. St. B. T. (1993). Mental models as an explanation of belief bias effects in syllogistic reasoning. *Cognition, 46*, 93-97.

Newstead, S. E. & Griggs, R. A. (1999). Premise Misinterpretation and Syllogistic Reasoning. *The Quarterly Journal of Experimental Psychology, 52A*, 1057-1075.

Newstead, S. E. (2000). Are there two different types of thinking? *Behavioural and Brain Sciences, 23*, 690-691.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual Differences in Deductive Reasoning. *The Quarterly Journal of Experimental Psychology, 57*, 33-60.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The Use of Statistical Heuristics in Everyday Inductive Reasoning. *Psychological Review, 90*, 339-363.

Nisbett, R. E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement.* Englewood Cliffs, N.J.: Prentice-Hall.

Noveck, I. A. & O'Brien, D. P. (1996). To What Extent Do Pragmatic Reasoning Schemas Affect Performance on Wason's Selection Task? *The Quarterly Journal of Experimental Psychology, 49A*, 463-489.

Novick, L. R. (1988). Analogical Transfer, Problem Similarity, and Expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 510-520.

Oakhill, J., Johnson-Laird, P.N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition. 31(2)*, 179-218.

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. (1996). Rational Explanation of the Selection Task. *Psychological Review, 103*, 381-391.

Over, D. E. (2003). Review. Who is Rational? Studies of Individual Differences in Reasoning. *Mind and Language, 18*, 225-230.

Pacini, R., Muir, F., & Epstein, S. (1998). Depressive Realism From the Perspective of Cognitive-Experiential Self-Theory. *Journal of Personality and Social Psychology, 74,* 1056-1068.

Pacini, R. & Epstein, S. (1999). The Relation of Rational and Experiential Information Processing Styles to Personality, Basic Beliefs, and the Ratio-Bias Phenomenon. *Journal of Personality and Social Psychology, 76,* 972-987.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond Abilities: A Dispositional Theory of Thinking. *Merrill-Palmer Quarterly, 39,* 1-21.

Perkins, D. N. & Grotzer, T. A. (1997). Teaching Intelligence. *American Psychologist, 52,* 1125-1133.

Phye, G. D. (1991). Advice and Feedback during Cognitive Training: Effects at Acquisition and Delayed Transfer. *Contempory Educational Psychology, 16,* 87-94.

Platt, R. D. & Griggs, R. A. (1993). Darwinian algorithms and the Wason selection task: A factorial analysis of social contract selection task problems. *Cognition, 48,* 163-192.

Pollard, P. & Evans, J.St.BT (1987). Content and context effects in reasoning. *American Journal of Psychology, 100 (1),* 41-60.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The Role of Analogy in Transfer Between Similar Problem States. *Cognitive Psychology,* 436-450.

Reeves, L. M. & Weisberg, R. W. (1993). Abstract Versus Concrete Information as the Basis for Transfer in Problem Solving: Comment on Fong and Nisbett (1991). *Journal of Experimental Psychology: General, 122,* 125-128.

Renner, C. H. & Renner, M. J. (2001). But I Thought I Knew That: Using Confidence Estimation as a Debiasing Technique to Improve Classroom Perfomance. *Applied Cognitive Psychology, 15,* 23-32.

Ross, B. H. & Kennedy, P. T. (1990). Generalizing From the Use of Earlier Examples in Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 42-55.

Royalty, J. (1995). Evaluating Knowledge-Based Statistical Reasoning. *Psychological Reports, 77,* 1323-1327.

Sa, W. C., West, R. F., & Stanovich, K. E. (1999). The Domain Specificity and Generality of Belief Bias: Searching for a Generalizable Critical Thinking Skill. *Journal of Educational Psychology, 91,* 497-510.

Sa, W. C. & Stanovich, K. E. (2001). The domain specificity and generality of mental contamination: Accuracy and projection in judgments of mental content. *British Journal of Psychology, 92,* 281-302.

Sedlmeier, P. & Gigerenzer, G. (1997). Intuitions About Sample Size: The Empirical Law of Large Numbers. *Journal of Behavioral Decision Making, 10,* 33-51.

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

Sloman, S.A. In: Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and Biases*. New York:. Cambridge University Press.

Smith, E. E., Langston, C. L., & Nisbett, R. E. (1992). The Case for Rules in Reasoning. *Cognitive Science, 16,* 1-40.

Speelman, C. P. & Kirsner, K. (1997). The Specificity of Skill Acquisition and Transfer. *Australian Journal of Psychology, 49,* 91-100.

Sperber, D. (1997). Intuitive and Reflective Beliefs. *Mind & Language, 12,* 67-83.

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

Sperber, D. & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language, 17,* 3-23.

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

Stanovich, K. E. & West, R. F. (1997). Reasoning Independently of Prior Belief and Individual Differences in Actively Open-Minded Thinking. *Journal of Educational Psychology, 89,* 342-357.

Stanovich, K. E. & West, R. F. (1998a). Cognitive Ability and Variation in Selection Task Performance. *Thinking and Reasoning, 4,* 193-230.

Stanovich, K. E. & West, R. F. (1998b). Individual Differences in Rational Thought. *Journal of Experimental Psychology: General, 127,* 161-188.

Stanovich, K. E. & West, R. F. (1999). Discrepancies Between Normative and Descriptive Models of Decision Making and the Understanding/Acceptance Principle. *Cognitive Psychology, 38,* 349-385.

Stanovich, K. E. & West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioural and Brain Sciences, 23,* 645-726.

Sternberg, R. J. & Davidson, J. E. (1987). Teaching Thinking to College Students: Some Lessons Learned from Experience. *Teaching Thinking and Problem Solving, 9,* 1-11.

Tomic, W. & Klauer, K. J. (1996). On the effects of training inductive reasoning: How far does it transfer and how long do the effects persist? *European Journal of Psychology of Education, 11,* 183-299.

Toplak, M. E. & Stanovich, K. E. (2002). The Domain Specificity and Generality of Disjunctive Reasoning: Searching for a Generalizable Critical Thinking Skill. *Journal of Educational Psychology, 94,* 197-209.

Tversky,A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Tversky,A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5,* 207-232.

Tversky, A. & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Confunction Fallacy in Probability Judgment. *Psychological Review, 90,* 293-315.

Volbrecht, V. J. & Schwartz, M. (1979). Transfer effects in a deductive reasoning problem. *Bulletin of the Psychonomic Society, 14,* 347-350.

Winne, P. H. (1996). A Metacognitive View of Individual Differences in Self-Regulated Learning. *Learning and Individual Differences, 8,* 327-353.

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

List of Appendices

Appendix A – Materials and Instructions Employed in Experiments 1 – 6

Appendix B – Statistical Tables of Statistical Analyses: t-tests and ANOVAS for Experiments 1 – 6

APPENDIX A1.1

Experiment 1: The Rational-Experiential Inventory (Pacini & Epstein, 1999) presented to all participants.

Following are a number of statements about feelings, beliefs, and behaviours. Read each statement and then circle the response from 1 to 5 that most applies to you. For example;

1	2	3	4	5
Definitely not true	Mostly not true	Equally true and false	Mostly true	Definitely true

1. I'm not that good at working out complicated problems.

2. Knowing the answer without having to understand the reasoning behind it is good enough for me.

3. I try to avoid situations that require thinking in depth about something.

4. I don't have a very good sense of intuition.

5. I am much better at working things out logically than most people.

6. I like to rely on my intuitive impressions.

7. I am not very good at solving problems that require careful logical analysis.

8. I enjoy intellectual challenges.

9. I usually have clear, explainable reasons for my decisions.

10. Using my "gut-feelings" usually works well for me in working out problems in my life.

11. Thinking hard and for a long time about something gives me little satisfaction.

12. Intuition can be a very useful way to solve problems.

13. Using logic usually works well for me in working out problems in my life.

14. I can usually feel when a person is right or wrong, even if I can't explain how I know.

15. I often go by my instincts when deciding on a course of action.

16. My snap judgements are probably not as good as most people's.

17. I enjoy thinking in abstract terms.

18. I don't like situations in which I have to rely on intuition.

19. I prefer complex to simple problems.

20. I suspect my hunches are inaccurate as often as they are accurate.

21. I think there are times when one should rely on one's intuition.

22. I don't like to have to do a lot of thinking.

23. I think it is foolish to make important decisions based on feelings.

24. I trust my initial feelings about people.

25. I don't reason well under pressure.

26. I don't think it is a good idea to rely on one's intuition for important decisions.

27. I have a logical mind.

28. I generally don't depend on my feelings to help me make decisions.

29. I enjoy solving problems that require hard thinking.

30. I would not want to depend on anyone who described himself or herself as intuitive.

31. I am not a very analytical thinker.

32. I tend to use my heart as a guide for my actions.

33. Learning new ways to think would be very appealing to me.

34. I believe in trusting my hunches.

35. I have no problem in thinking things through carefully.

36. When it comes to trusting people, I can usually rely on my gut feelings.

37. Thinking is not my idea of an enjoyable activity.

38. Reasoning things out carefully is not one of my strong points.

39. If I were to rely on my gut feelings, I would often make mistakes.

40. I hardly ever go wrong when I listen to my deepest "gut-feelings" to find an answer.

APPENDIX A1.2

Experiment 1: The Thinking Dispositions Questionnaire (Stanovich & West, 1998a) presented to all participants.

Following are a number of statements about various topics. Read each statement and then circle the response from 1 to 5 that most applies to you. For example;

1	2	3	4	5
Strongly Disagree	Slightly Disagree	Neither Agree/Disagree	Slightly Agree	Strongly Agree

1. It is better to simply believe in a religion than to be confused by doubts about it.
2. Even if my environment (family, neighbourhood, schools) had been different, I probably would have the same religious views.
3. Opening an umbrella indoors will increase one's chances of misfortune in the near future.
4. A professor's job is to communicate the facts to his or her students.
5. Right and wrong never change.
6. Changing your mind is a sign of weakness.
7. Difficulties can usually be overcome by thinking about the problem, rather than through waiting for good fortune.
8. It is bad luck to have a black cat cross your path.
9. Intuition is the best guide in making decisions.
10. Considering too many different opinions often leads to bad decisions.
11. It is advisable to consult your horoscope daily.
12. I can't enjoy the company of people who don't share my moral values.
13. People should always take into consideration evidence that goes against their beliefs.
14. Once a person decides on an occupation, his or her professional behaviour is mostly set.
15. Coming to decisions quickly is a sign of wisdom.
16. There is nothing wrong with being undecided about many issues.

17. My beliefs would not have been very different if I had been raised by a different set of pàrents.
18. The number 13 is unlucky.
19. Basically, I know everything I need to know about the important things in life.
20. Good teachers never let you leave the classroom with doubts about the subject matter.
21. If I think longer about a problem I will be more likely to solve it.
22. A person of good character usually does what he or she is told to do.
23. There are two kinds of people in this world: those who are for the truth and those who are against the truth.
24. Of all the different philosophies, which exist in the world, there is probably only one which is correct.
25. In most situations requiring a decision, it is better to listen to someone who knows what they are doing.
26. The best courses emphasise practical rather than theoretical matters.
27. Astrology can be useful in making personality judgements.
28. A person should always consider new possibilities.
29. Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups.
30. I have personal possessions that bring me luck at times.

APPENDIX A1.3

Experiment 1: The Argument Evaluation Test (Stanovich & West, 1997) modified for a British population and presented to all participants.
Part One consists of 25 items to which participants indicate their degree of belief.

Opinions

Please indicate your degree of agreement or disagreement with the following beliefs. Answer each item in order and do not return to items. There is no right or wrong answer, just go with your first impressions.

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

1. It is more dangerous to travel by air than by car.
2. Seventeen year olds should have the legal right to drink alcoholic beverages.
3. The relative ease with which handguns can be obtained has resulted in an unnecessarily high murder rate in America.
4. Interviews should be given a higher weighting in university admissions than A levels or Access courses.
5. The present state pension scheme is unfair to people who are now retired.
6. Computers cannot think.
7. The social security system should be drastically reduced in size.
8. Women should stay home and take care of the children while they are young.
9. Judges should sentence more young offenders to prison for their crimes.
10. Taxes in the England are too high.

11. The speed limit should be 80mph on the motorways.

11. Seatbelts should always be worn when travelling by car.

13. The tax on petrol should not be raised significantly.
14. The death penalty should be brought back as an ultimate punishment.
15. Students should have a stronger voice than the general public in setting university policies.

16. The legal voting age should not be raised from 18 years to 21 years.

17. The national debt should be reduced by cutting MP's salaries.

18. Today's footballers are grossly overpaid.

19. Labour unions should be eradicated because they are a major cause of the downfall of the British economy.

20. Capital punishment should be outlawed in America.

21. Smoking should be banned in all enclosed public places.

22. It is unfair for a new insurance policy holder to collect a huge insurance payment soon after obtaining the insurance policy.

23. Children who play computer games will become anti-social adults.

24. The government's plan to get single parents back to work should be continued.

25. The governments plan to give £1000 to every new born baby is a waste of financial resources.

APPENDIX A1.3 Cont.

Part Two. Participants are then required to evaluate the arguments in a second testing session.

Evaluation of Counter-Arguments Questionnaire

Instructions: We are interested in your ability to evaluate counter-arguments. First, you will be presented with a belief held by an individual named James. Following this, you will be presented with James's premise or justification for holding this particular belief. A Critic will then offer a counter-argument to James's justification for the belief. (Assume that the Critic's statement is factually correct.) Finally, James will offer a rebuttal to the Critic's counter-argument. (Assume that James's rebuttal is also factually correct.) You are to evaluate the strength of James's rebuttal to the Critic's counter-argument, regardless of your feeling about the original belief or James's premise.

James's belief: It is more dangerous to travel by air than by car.

James's premise or justification for belief: It is more dangerous to travel by air than by car because air accidents are more likely to involve fatalities.

Critic's counter-argument: Passengers are 3 times more likely to be killed per mile travelled in a car as compared to a plane (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Because reckless or drunk drivers cause the great majority of all automobile accidents (assume statement factually correct), car travel is at least safer than air travel for people who wear safety belts and travel with sober and careful drivers.

Indicate the strength of James's rebuttal to the Critic's counter-argument:

Very Weak = 1   *Weak = 2*   *Neither = 3*   *Strong = 4*   *Very Strong = 5*

James's belief: Seventeen year olds should have the legal right to drink alcoholic beverages.

James's premise or justification for belief: Seventeen year olds are just as responsible as eighteen year olds, so they ought to be granted the same drinking rights as other adults.

Critic's counter-argument: Seventeen year olds are 3 times more likely to be involved in an automobile accident while under the influence of alcohol than eighteen year olds (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Seventeen year olds will drink no matter what the law says (assume statement factually correct), so it is useless to try to legislate that they not drink.

James's belief: The relative ease with which handguns can be obtained has resulted in an unnecessarily high murder rate in America.

James's premise or justification for belief: Countries that strictly limit handgun ownership have far fewer murders.

Critic's counter-argument: People without handguns will use knives or other weapons instead (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Since handguns are particularly lethal and most violent attacks are impulsive acts (assume statement factually correct), attacks with handguns are more likely to result in death.

James's belief: Interviews should be given a higher weighting in university admissions than A levels or Access courses.

James's premise or justification for belief: Weighting interviews higher than A levels or Access courses will result in better people being admitted to universities.

Critic's counter-argument: Both A levels and Access courses are much better predictors of success in universities than interviews (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Interviews sometimes uncover important personal information that would be missed by an impersonal process that relies heavily on grades and courses (assume statement factually correct).

James's belief: The present state pension scheme is unfair to people who are now retired.

James's premise or justification for belief: The present state pension scheme is unfair to retired people because these people get far too little money.

Critic's counter-argument: Currently, pensioners are drawing from the pension scheme over four times what they contributed to the system before they die (assume statement factually correct).

James's rebuttal to Critic's counter-argument: These figures don't take into account the fact that pensioners are being paid in pounds that have been greatly eroded over the years due to inflation (assume statement factually correct).

James's belief: Computers cannot think.

James's premise or justification for belief: Computers cannot think because computers do not produce anything that is original (i.e. anything that human's haven't already produced).

Critic's counter-argument: A computer has produced a proof of one of Euclid's geometry theorems that no human had produced before (assume statement factually correct).

James's rebuttal to Critic's counter-argument: The computer that produced that proof was built by humans (assume statement factually correct).

James's belief: The social security system should be drastically reduced in size.

James's premise or justification for belief: The social security system should be drastically reduced in size because social security recipients take advantage of the system and buy unnecessary luxuries with their dole money.

Critic's counter-argument: Ninety-five percent of social security recipients use their dole money to obtain the bare essentials for their families (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Many people who are on social security are lazy and don't want to work for a living (assume statement factually correct).

James's belief: Women should stay at home and take care of the children while they are young.

James's premise or justification for belief: Only a mother can provide the quality of care young children both need and deserve.

Critic's counter-argument: Women who are in self-fulfilling careers are confident parents who spend as much quality time with their young children as nonworking women (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Most women work out of necessity, not because they find their employment fulfilling (assume statement factually correct).

James's belief: Judges should sentence more young offenders to prison for their crimes.

James's premise or justification for belief: More young offenders should be sentenced to prison because severe punishment will act as a deterrent to future criminal activities.

Critic's counter-argument: Youths who are sent to prison for relatively minor infractions have a greater likelihood of becoming adult criminals than youths who are placed on parole in a community rehabilitation programme (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Youths who commit even relatively minor infractions are more likely to engage in criminal activities as adults than young non-offenders (assume statement factually correct).

James's belief: Taxes in England are too high.

James's premise or justification for belief: British industry is at a competitive disadvantage because our taxes are too high.

Critic's counter-argument: Taxes in England are among the very lowest in the industrial world (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Politicians waste millions of dollars on useless pet projects (assume statement factually correct).

James's belief: The speed limit should be 80mph on the motorways.

James's premise or justification for belief: Because the motorway system was designed for high speed traffic, it is safe to drive at 80mph.

Critic's counter-argument: You are more likely to die as a result of an accident if you are travelling at 80mph than you are if you are travelling at 70mph (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Many people will drive at 80mph whether or not the legal limit is 70 (assume statement factually correct).

<u>James's belief:</u> Seatbelts should always be worn when travelling in a car.

James's premise or justification for belief: Seatbelts should always be worn to make travelling by car safer.

Critic's counter-argument: There are times when your life may be saved by your being thrown free of a car during an accident (assume statement factually correct).

James's rebuttal to Critic's counter-argument: You are several times more likely to be killed if you are thrown from a car (assume statement factually correct).

James's belief: The tax on petrol should not be raised significantly.

James's premise or justification for belief: The tax on petrol should not be raised significantly because the economy will be hurt by higher taxes.

Critic's counter-argument: Taxes could always be lowered in other areas to compensate for the increased tax in petrol, and raising the cost of petrol will encourage conservation which would be good for the economy in the long run (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Increased tax on petrol would hit lower and middle income people harder than wealthier people (assume statement factually correct).

James's belief: The death penalty should be brought back as an ultimate punishment.

James's premise or justification for belief: The judicial system should reinstate the death penalty because fear of capital punishment will serve as a strong deterrent for potential murderers.

Critic's counter-argument: Evidence strongly suggests that murderers either kill impulsively or else they assume that they will never get caught (assume statement factually correct).

James's rebuttal to Critic's counter-argument: If murderers get the death penalty, they get what they deserve (assume statement factually correct).

James's belief: Students should have a stronger voice than the general public in setting university policies.

James's premise or justification for belief: Because students are the ones who must ultimately pay the costs of running the university through tuition, they should have a stronger voice in setting university policies.

Critic's counter-argument: Tuition covers less than one half the cost of an education at most universities (assume statement factually correct), so the taxpayers should have a stronger say in the policies.

James's rebuttal to Critic's counter-argument: Because it is the students who are directly influenced by university policies (assume statement factually correct), they are the ones who should have the stronger voice.

James's belief: The legal voting age should not be raised from 18 years to 21 years.

James's premise or justification for belief: The legal voting age should not be raised because, when it comes to voting, 18 year olds are as responsible as 21 year olds.

Critic's counter-argument: Eighteen year olds do not generally take their voting responsibilities seriously, since they rarely vote and have been repeatedly demonstrated to have little knowledge of current events (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Those eighteen year olds who do vote are relatively will informed and take their responsibilities seriously (assume statement factually correct).

James's belief: The national debt should be reduced by cutting MP's salaries.

James's premise or justification for belief: MP's salaries are very high, and cutting them would make a significant step towards paying off the huge national debt.

Critic's counter-argument: The national debt is so large that totally eliminating all MP's salaries would still hardly make a dent in it (assume statement factually correct).

James's rebuttal to Critic's counter-argument: Members of Parliament, whose actions are to a considerable extent responsible for the huge national debt, earn salaries several times higher than the national average (assume statement factually correct).

James's belief: Todays footballers are grossly overpaid.

James's premise or justification for belief: It is unfair for fans to have to pay the high ticket prices required to meet the extremely high salary demands of footballers.

Critic's counter-argument: Because many fans will only support teams with celebrities and winning records (assume statement factually correct), footballers earn their money.

James's rebuttal to Critic's counter-argument: The footballers of yesteryear are at least as good as those of today, yet they didn't earn millions of pounds a year (assume statement factually correct).

James's belief: Labour unions should be eradicated because they are a major cause of the downfall of the British economy.

James's premise or justification for belief: Labour unions are a major cause of the downfall of the British economy because they force management to pay extravagantly high wages, which corporations simply cannot bear if they intend to remain competitive on the world market.

Critic's counter-argument: Labour unions keep management from abusing and taking advantage of workers (assume statement factually correct), so they are necessary.

James's rebuttal to Critic's counter-argument: Most of the abuses that took place in the past are now prevented by law (assume statement factually correct); thus, unions are no longer needed for that purpose.

<u>James's belief:</u> Capital punishment should be outlawed in America.

<u>James's premise or justification for belief</u>: Capital punishment should be outlawed because killing is wrong and the moral costs of sentencing an innocent person to death are too great.

<u>Critic's counter-argument</u>: The prison system is very overcrowded, and it costs the country over $25,000 per prisoner each year to maintain each prisoner (assume statement factually correct).

<u>James's rebuttal to Critic's counter-argument</u>: The cost to the state of processing a capital punishment case through to completion averages about ten-million dollars in court and legal costs for each case (assume statement factually correct).

<u>James's belief:</u> Smoking should be banned in all enclosed public places.

<u>James's premise or justification for belief</u>: Smoking should be banned in all enclosed public places because even second hand smoke poses a significant health risk to non-smokers.

<u>Critic's counter-argument</u>: Since many smokers already refrain from smoking in places where their second hand smoke poses a health risk to others (assume statement factually correct), it is unnecessary to severely restrict smoking locations.

<u>James's rebuttal to Critic's counter-argument</u>: While it may be true that many smokers are considerate, it is equally true that many smokers are not so considerate (assume statement factually correct). Banning smoking would be an effective way to ensure that many of us won't be subjected to the risks posed by second hand smoke

---

<u>James's belief:</u> It is unfair for a new insurance policy holder to collect a huge insurance payment soon after obtaining the insurance policy.

<u>James's premise or justification for belief</u>: It is unfair for a new insurance policy holder to collect a huge insurance payment soon after obtaining the insurance policy because the total amount they would have contributed to the insurance fund would not come close to covering the cost of the payment.

<u>Critic's counter-argument</u>: Because only a small fraction of all new insurance policy holders have a need to collect on their policies, the large sums paid to the few new policy holders who need payment are covered by the insurance payments of the remaining new policy holders (assume statement factually correct).

<u>James's rebuttal to Critic's counter-argument</u>: Individuals who have been insured with a company for a long period of time have contributed much more to the insurance fund than new policy holders (assume statement factually correct).

<u>James's belief:</u> Children who play computer games will become anti-social adults.

<u>James's premise or justification for belief</u>: Children who play computer games will become anti-social adults because they won't have learned the social skills necessary for them to interact with other adults.

<u>Critic's counter-argument</u>: Computer games are a modern trend and are no different socially from reading a book (assume statement factually correct).

<u>James's rebuttal to Critic's counter-argument</u>: Eighty percent of children who play computer games spend at least 3 hrs a day on them, whilst children who prefer to read, read for much less than 3 hours a day (assume statement factually correct).

<u>James's belief:</u> The government's plan to get single parents back to work should be continued.

<u>James's premise or justification for belief</u>: Getting single parents back to work will give them a sense of self-worth and an aim in life.

<u>Critic's counter-argument</u>: Most working people still have no sense of self-worth (assume statement factually correct).

<u>James's rebuttal to Critic's counter-argument</u>: Seventy-five percent of single parent families that have returned to work, have reported being much happier in their home life (assume statement factually correct),.

<u>James's belief:</u> The government's plan to give £1000 to every new born baby is a waste of financial resources.

<u>James's premise or justification for belief</u>: The government's plan to give £1000 to every new born baby is a waste of financial resources because it will encourage people to have babies for the wrong reasons.

<u>Critic's counter-argument</u>: Most parents say that they will use the money to invest in their child's future education (assume statement factually correct).

<u>James's rebuttal to Critic's counter-argument</u>: The government would get a better return investing all the money into education now (assume statement factually correct),.

APPENDIX A1.4

Experiment 1: Four full-time members of staff at the University of Plymouth, all highly qualified in the field of human reasoning and decision making, one philosopher and the principle researcher rated the arguments on the AET. The table below displays the correlations between the six expert raters. The median correlation between the six experts was 0.65.

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
Rater 1					
Rater 2	.70***				
Rater 3	.75***	.54***			
Rater 4	.65***	.79***	.58***		
Rater 5	.75***	.64***	.63***	.75***	
Rater 6	.71***	.72***	.64***	.63***	.62***

***p<.001 (two-tailed)

Table of correlations of the ratings of the arguments on the AET

APPENDIX A1.5

Experiment 1: Instructions followed by the nine Law of Large Numbers problems, three belief-consistent, three belief-inconsistent and three belief-neutral (Klaczynski, Gordon, & Fauth, 1997)

Instructions

On the following pages you will find nine arguments based on different professions. Please read through them carefully and answer the questions which follow each one. Attempt them in the order in which they are presented and when finished do not turn back.

Following each argument could you please rate the strength of the conclusion (1 = extremely weak; 9 = extremely strong) and how persuaded you are by the argument (1 = extremely unconvinced; 9 = extremely convinced). Space is also provided under each argument for you to write an explanation of why you believe the conclusion to be weak or strong and the argument to be unconvincing or convincing.

If you have any questions, please ask the experimenter.

Belief-Consistent

Structure 1:

Karen is a secretary for Dr. T., a respected custodian at his office. During the time she has worked for him, Dr. T. has always been polite, never told any "dirty" jokes, never made any "passes" at her, and never done anything that even looked like sexual harassment. All of his behaviors have made Karen very comfortable; she is hoping to keep her current job as long as possible. Her friend Helen is thinking about taking a job as secretary for a different custodian. In giving Helen advice about whether to take the job, Karen said the following:

Karen: I've never been so comfortable in my life! He just never does anything wrong or anything that makes me feel uncomfortable. He never pulls any of that "male-macho" stuff. I'd love to keep working in his office forever, but if I can't, you can bet I'll look for another job with a custodian. Dr. T. is so NON-sexist, that it's unbelieveable! Before you make any hasty decisions about working in the office of anybody other than a custodian, I'd think carefully. If you want a bunch of men constantly touching you, hitting on you, and groping you, then take those jobs. Look: I've had experience with a custodian and it was great. I really recommend it.

Based on the evidence used, how weak or strong do you think the conclusion is?

After reading the argument, how convinced are you?

Use the space below to answer the following question. Please comment on the conclusion. In no more than 2 or 3 sentences, explain <u>why</u> you think the conclusion is good or bad:

Structure 2:

Dr. S. Hernstein is psychologist who has spend several years determining whether people in certain occupations make better parents than people in other occupations. In a recent study, he compared approximately 500 custodians to approximately 500 engineers. In the report he published, Dr. Hernstein found that:

"It is clear from these results that custodians make somewhat worse parents than engineers. Custodians physically abuse their children, pay less attention to their children's schoolwork, and participate in fewer activities than engineers. Also, custodians, on the average, report that they have less fun with their children than engineers and are more likely to wish that they had never had any children than engineers."

In response to his paper, Dr. R. Thomas, also a respected psychologist, wrote a paper in which he attempted to discredit Dr. Hernstein's argument. Dr. Thomas wrote:

"I have difficulty believing Dr. Hernstein's study for two reasons. First, in my laboratory, I work with a custodian who loves his children very much. Second, talking about the 'average' behavior of custodians doesn't tell me anything about how any one custodian will act. On this basis, I feel we can effectively dismiss Hernstein's argument."

Structure 3:

Ed G. did his thesis in psychology on the subject of conformity to authority in different occupational groups. In his study, 3 engineers, 3 custodians, and 3 local politicians participated. Below is a summary of Ed's study.

When they came into the psychology laboratory, Ed commanded each subject to do several menial tasks, such as grading undergraduate papers, making copies of class assignments, and taking notes during a meaningless lecture. As signs of conformity, Ed measured the following: How quickly the subjects responded to his demands, whether they ever questioned his orders, and whether they actually followed his demands. He then constructed a scale like that below; the results of his study are to the right of the scale:

	Results
1 = Very conforming;	1 engineer, 2 politicians
2 = Somewhat conforming;	1 engineer,
3 = Moderate conforming;	1 engineer, 1 politician
4 = Somewhat Nonconforming;	1 custodian
5 = Very Nonconforming	2 custodians

Because these data showed that custodians were either very nonconforming or somewhat nonconforming and because the other two groups usually obeyed his orders, Ed concluded that custodians are much less conforming than either engineers or politicians.

Belief-Inconsistent

Structure 1:

An editorial in a local newspaper recently criticised the profession of being a custodian. The journalist's argument was:

I've got a friend who's a custodian, and I wouldn't want my children to be influenced by him. I went to visit him at work the other day, just to say "Hi." What did I find? The custodian wasn't exactly working--he was in the lounge taking a nap! Friends of mine tell me he's always sleeping or loafing or messing around! What am I supposed to think? My conclusion: I don't think there's a lazier group of people than custodians!

Structure 2:

Lisa and Roy are having an argument over whether custodians are more likely to get divorced than other people. Roy claims the marriages of custodians are much more satisfying than those of other people, but Lisa believes that custodians' marriages are very unhappy and frequently end in divorce.

Roy:
I've been working as a lawyer for over 15 years, and handle people's divorces all the time. During all that time, I've seen hundreds of couples in bad marriages that wind up in divorce and they've been from just about every occupation you can imagine, but by far fewer of those divorces have involved custodians than any other occupation. I've seen enough to know that a marriage is pretty likely be happy, satisfying, and to stay together when a custodian's involved.

Lisa:
You're wrong and you know it! Two of my sisters are custodians, remember? I talk to them every day and they're always talking about problems with their marriages. In fact, I know several people who are custodians and they have the least satisfying marriages I've know of. I've seen it for myself, you haven't! If you've seen the custodians I've seen, it's pretty easy to figure out that custodians are a good bet to be unhappy in their marriages.

Structure 3:

John works for a London testing company. His company has been hired to measure the IQs or intelligence of Ministry of Defence employees. John himself has been asked to give an IQ test to 10 people who are MOD employees. Two people were secretaries, two were doctors, 2 were custodians, 2 were architects, and 2 were dentists. He uses the best IQ test available, one that is respected across the country. After he had tested each person once, he put his findings in a table:

Person:	IQ:
Secretaries	98 and 101
Custodians	90 and 97
Doctors	115 and 117
Architects	106 and 111
Dentists	102 and 108

Because a person with average intelligence should have an IQ of around 100 and because IQs below are low, especially for most professions, John concludes that custodians are much less intelligent than the other groups of people he tested.

Belief-Neutral

Structure 1:

Amy is thinking about being a chemist, and her friend Jill is trying to talk her into it. Jill's argument is, "Look, my father was a chemist, and after 20 years of being a chemist, he's still energetic, trying to do his best, and is very happy. It's a great job to do. You can just take one look at my dad and you'll know that he's got a great job."

Structure 2:

Ken and Toni are arguing over whether chemists have more psychological problems than other people.

Ken's argument is, "Of course chemists have fewer problems! All that academic and school work keeps their minds in good psychological health! Remember the chemists that we've known? Mr. Cross was the happiest guy I knew and never had any problems and Mrs. Davis was always helping other people solve their own problems! How can you argue with that?"

Toni's argument is, "I don't think you're right. I just read a newspaper article that compared hundreds of chemists with hundreds of other people. That study showed that chemists are much more likely to have psychological problems than other people."

Structure 3:

Fran works for an employment agency. As part of her job, she has to interview people from a huge number of professions. Recently, she interviewed a young chemist who was looking for a job in the county. During the interview, the chemist was very calm, had complete control of herself, and expressed herself very clearly. Fran was very impressed. In fact, it was the best interview Fran had ever seen. From this experience, Fran has concluded that chemists have great self-confidence and that in their training as chemists they learn how to stay cool under pressure.

The rating scales following each problem:

Based on the evidence used, how weak or strong do you think the conclusion is?

1	2	3	4	5	6	7	8	9
Extremely Weak		Somewhat Weak		Average: Neither Weak nor Strong		Somewhat Strong		Extremely Strong

After reading the argument, how convinced are you?

1	2	3	4	5	6	7	8	9
Extremely Unconvinced		Somewhat Unconvinced		Not Sure		Somewhat Convinced		Extremely Convinced

Use the space below to answer the following question. Please comment on the conclusion. In no more than 2 or 3 sentences, explain <u>why</u> you think the conclusion is good or bad:

____

____

____

____

____

APPENDIX A1.6

Experiment 1: Instructions followed by the nine Experiment Evaluation problems, three belief-consistent, three belief-inconsistent and three belief-neutral (Klaczynski, Gordon, & Fauth, 1997)

Instructions

On the following pages you will be presented with several summaries of actual psychological and sociological research. The research involves participants from several dozen occupations, one of which may be your intended occupation. Many of the findings that pertained to other occupations have been deleted from the summaries and only information relevant to your occupation and a few other randomly selected occupations will be presented.

Following each problem could you please rate the strength of the researcher's conclusion (1 = extremely weak; 9 = extremely strong) and the validity of the experiment i.e. does the experiment measure what the researcher suggests it measures (1 = extremely invalid; 9 = extremely valid). Space is also provided under each problem for you to write an explanation of why you believe the researcher's conclusion to be weak or strong and the experiment to be valid or invalid.

Belief-Consistent

Structure 1:

A famous sociologist conducted a study looking at the effects of a person's occupation on their stress levels. He was particularly interested in the stress factors that may be more apparent in the workplace for some professions over others. Large groups of accountants, engineers and nurses were recruited to take part in the study. Each occupational group was asked to attend the lab for an afternoon of testing on different days. Once at the lab, the participants were split into smaller groups of eight. Video-taped focus groups were led by experienced research assistants who introduced stress inducing topics for discussion. The groups consisting of engineers and nurses were led into discussion about salaries, shiftwork and management issues. The groups consisting of accountants were led into discussion about salaries and holidays. After the focus group which lasted half an hour, participants were required to complete a self-report inventory to ascertain their stress levels. The results showed that accountants had lower self-reported stress levels than either nurses or engineers after the focus groups and on watching the videotapes the sociologist concluded that the focus groups consisting of accountants included less stress-filled discussion. His reported findings indicate that accountancy as an occupation leads to lower stress levels than either engineering or nursing.

Structure 2:

Professor A. recently ran a large scale study looking at positive and negative personality traits in individuals from different occupations. Positive personality traits observed for example were cheerfulness, patience, honesty and open-mindedness, whereas examples of negative personality traits observed were distrust of others, pessimism, overbearing behaviours and laziness. With the aid of trained researchers, seventy-five accountants randomly selected from all over the country were followed and observed over a period of six months going about their work and any personality traits were monitored. Seventy-five engineers were also observed for six months but as their work generally involves a great deal of travelling, they were pre-selected from half a dozen factories in the Midlands. Along with being observed, participants were requested to complete a monthly questionnaire, designed to elicit certain responses dependent on the individual's personality. At the end of six months it was found that the researchers had monitored more positive personality traits than negative ones in accountants whereas the reverse was found for engineers. Also, the completion rate for the self-report personality scale was significantly higher for accountants than engineers and more positive responses were made by the accountants. Professor A. concluded from these results that accountants have more positive personality traits than engineers.

Structure 3:

A leading psychologist from a University Hospital conducted a study to look at overall satisfaction with life in individuals from different occupations. For the research she compared teachers, nurses, secretaries, accountants and waiters/waitresses. Sixty people were recruited from each occupation and for a fortnight they were each required to note down in a diary each positive thought that they had, day or night. At the end of the study period the results were collated and it was found that overall teachers had an average of 7 positive thoughts a day, nurses and waiters/waitresses had an average of 6 positive thoughts a day, secretaries had an average of 4.3 positive thoughts a day and accountants had an average of 11.8 positive thoughts a day. From these results the psychologist noted that accountants had significantly more positive thoughts each day than teachers, nurses, secretaries or waiters/waitresses. She concluded that this indicated that accountants were overall much more satisfied with their lives than individuals from any of the other occupations.

Belief-Inconsistent

Structure 1:

Dr. H. is a researcher who is interested in finding out how capable people in different occupations are of holding pleasant conversations with strangers. Dr. H. is especially interested in finding out how skilled accountants, teachers, and lawyers are in dealing with strangers because people in these professions must deal with strangers on a daily basis. To observe their conversations, Dr. H. watched a large number of teachers and lawyers in conversations with Amy, an attractive young female who is one of Dr. H's laboratory assistants. Next, Dr. H. observed a large number of accountants having a conversation with Jim, another one of Dr. H's laboratory assistants. The results of her study showed that teachers and lawyers had longer and more pleasant conversations and had much more eye contact than the accountants. On the basis of these findings, Dr. H. concluded that accountants are less socially skilled and less capable of having conversations with strangers than either teachers or lawyers.

Structure 2:

Two well-known researchers have conducted several studies on the relationship between one's occupation and the tendency to conform to authority. The occupations of being a teacher and being an accountant were compared in one study. In this study, 50 accountants were contacted at local businesses and agreed to be in the study. At the time of the study, teachers were hard to locate and contact. Thus, the researchers recruited 50 teachers from a conference on leadership that was taking place in a nearby city. Next, each person was brought to the researchers' laboratory. At the laboratory, both the accountants and the teachers were given several orders to complete menial tasks. For example, they were ordered to grade the papers of several dozen undergraduates. At the completion of the experiment, the researchers found that the accountants were far more likely to obey the commands than were the teachers. The teachers were much more likely to question the commands and to refuse. The researchers concluded that the experience of being a teacher results in a greater sense of independence and of inner strength than does being an accountant.

Structure 3:

Dr. Bill R. is a psychologist interested in determining whether sexual harassment is more likely to occur in some occupations than in others. To conduct his research, he included in his study accountants, architects, biologists, and salespersons. In each occupational group, he asked 80 people to be in the study. To measure sexual harassment, Bill observed people in each group at work and counted the number of times each person told jokes with sexual content. At the end of his study, Bill found that the average accountant told 6.5 sexual jokes per month. Members of the other occupations, on the other hand, told an average of only 2.0 sexual jokes per month. Therefore, accountants told more than three times as many sexual jokes than people in other occupations. Based on this, Dr. R. concluded that accountants are involved in more sexual harassment than are people who are members of other occupations.

Belief-Neutral

Structure 1:

Professor Y. wanted to find out whether individuals who worked in any one occupation were more or less socially skilled than individuals who worked in another. She randomly selected 40 dentists, 40 doctors and 40 teachers from all over the United Kingdom and invited them to a conference and dinner together. The teachers couldn't make it on the same day therefore their conference/dinner was held two weeks later. At the conferences and dinners, individuals were watched by trained researchers to observe their behaviours. For example, introducing themselves and introducing others. On collation of all the data, Professor Y. found that dentists showed the most social behaviours, closely followed by doctors, but with significantly less social behaviours were the teachers. In reporting her findings, Professor Y. concluded that dentists are more socially skilled individuals than doctors or teachers.

Structure 2:

Dr. S. and Dr. A. designed a study to ascertain whether there were differences in how people think dependent on their occupation. They were particularly interested in how biased people from different occupations were. Forty nurses and engineers were randomly selected from throughout Devon and Cornwall to participate in the study. Forty dentists were also recruited from an 'holistic approach to patient care' conference being held locally. The study took 45-60 minutes to complete and participants were required to read 20 scenarios depicting problems faced by fictitious people. They then had to rate on a scale of 1-10 whether they thought the problem was caused by the person involved themselves or by circumstances, and then describe ways that the problems could be solved and by whom. Dr. S. and Dr. A. found that the nurses and engineers were more likely to blame the fictitious individual for their problem and more likely to present only one solution to the problem – usually the individual having to change something about themselves. Dentists on the other hand would tackle the problem from different angles and offer many solutions. They concluded in their report that dentists are less 'person-blaming and more open to considering alternatives than either nurses or engineers.

Structure 3:

Dr. Wendy P. had a very strong interest in group work and specifically group cohesion. She wanted to find out whether people who worked in certain occupations were more able to work together as a group when tackling a problem than people from other occupations. Dentists, teachers and electricians were included in the study, 20 in each group. Each group was taken into the lab at different times and split into smaller groups of four. Each group of 4 was then given a task to complete in 30 minutes which was to design a recruitment poster advertising their profession to school leavers. She found that dentists overall produced the most eye-catching a informative posters. Dr. Wendy P. concluded from these results that dentists are more able to work as a group, and therefore exhibit more group cohesion than either teachers or electricians.

The rating scales following each problem:

Could you please rate the strength of the researchers conclusion on the following scale.

1	2	3	4	5	6	7	8	9
Extremely Weak		Somewhat Weak		Average: Neither Weak nor Strong		Somewhat Strong		Extremely Strong

Could you please rate the validity of the experiment on the following scale.

1	2	3	4	5	6	7	8	9
Extremely Invalid		Somewhat Invalid		Not Sure		Somewhat Valid		Extremely Valid

Use the space below to answer the following question.: Based on the experiment above please could you explain why you think the researcher's conclusion was weak/strong and why you think the experiment was valid/invalid.

APPENDIX A1.7

Experiment 1: The sixteen syllogistic reasoning problems, eight involving definitional content and eight consisting of empirical content.

Definitional:

No dogs are unhappy,
Some mammals are unhappy,
Therefore,
Some mammals are not dogs.

No birds are loving,
Some robins are loving,
Therefore,
Some robins are not birds.

No reptiles are fun-loving,
Some lizards are fun-loving,
Therefore,
Some reptiles are not lizards.

No trout are grumpy,
Some fish are grumpy,
Therefore,
Some trout are not fish.

Some mice are timid,
No tigers are timid,
Therefore,
Some mice are not tigers.

Some sheep are ferocious,
No four-legged animals are ferocious,
Therefore,
Some sheep are not four-legged animals.

Some alsatians are shy,
No dogs are shy,
Therefore,
Some dogs are not alsatians.

Some crustaceans are mean-tempered,
No crabs are mean-tempered,
Therefore,
Some crabs are not crustaceans.

Empirical:

No police dogs are vicious,
Some highly trained dogs are vicious,
Therefore,
Some highly trained dogs are not police dogs.

No nutritional things are inexpensive,
Some vitamins are inexpensive,
Therefore,
Some vitamin tablets are not nutritional.

No addictive things are inexpensive,
Some cigarettes are inexpensive,
Therefore,
Some addictive things are not cigarettes.

No millionaires are hard workers,
Some rich people are hard workers,
Therefore,
Some millionaires are not rich people.

Some priests are healthy,
No religious people are healthy,
Therefore,
Some religious people are not priests.

Some healthy people are unhappy,
No astronauts are unhappy,
Therefore,
Some astronauts are not healthy people.

Some good swimmers are smokers,
No deep sea divers are smokers,
Therefore,
Some good swimmers are not deep sea divers.

Some judges are accountants,
No well educated people are accountants,
Therefore,
Some judges are not well educated people.

APPENDIX A1.8

Experiment 1: Four selection task problems, two arbitrary and two deontic versions.

Arbitrary Versions

PROBLEM 1: The Wason Selection Task

Each of the boxes below represents a card lying on a table. Each one of the cards has a letter on one side and a number on the other side.

The rule is: "If a card has an A on its letter side, then it has an 8 on its number side."

As you can see, two of the cards are letter-side up, and two of the cards are number-side up. Your task is to decide which card or cards must be turned over in order to find out whether the rule is true or false.

Please indicate your choice or choices by circling either the 'a' (turn) or 'b' (not turn) below each card.

A	8	3	D

PROBLEM 2: The Destination Problem

Each of the tickets below has a destination on one side and a mode of travel on the other side.

The rule is: "If 'Glasgow' is on one side of the ticket, then 'train' is on the other side of the ticket."

As you can see, two of the tickets are destination-side up, and two of the tickets are mode of travel-side up. Your task is to decide which ticket or tickets you would need to turn over in order to find out whether the rule is true or false.

Please indicate your choice or choices by circling either the 'a' (turn) or 'b' (not turn) below each ticket.

Glasgow	Edinburgh	Train	Coach

Deontic Versions

PROBLEM 3: The Sears (Debenhams) Problem

Suppose that you are the assistant manager at Debenhams, and it is your job to check sales receipts to make sure they are properly filled out according to a rule.

The rule is: "Any sale over £30 must be approved by the section manager, Mr. Jones. "

The amount of the sale is on one side of each receipt, and the space for the approval signature is on the other side. Which of the sales receipts shown below would you need to turn over in order to find out whether or not the rule is being violated?

Please indicate your choice or choices by circling either the 'a' (turn) or 'b' (not turn) below each sales receipt.

£70	£22	Approved	Not Approved

PROBLEM 2: The Drinking Age Problem

Imagine that you are a police officer on duty, walking through a local bar. It is your job to ensure that the drinking laws are in effect in this bar. When you see a person engaging in certain activities, the laws specify that certain conditions must first be met.

One such law is "If a person is drinking beer then the person must be over 18 years of age."

Each of the boxes below represents a card lying on a table. There are two pieces of information about a person on each card. Whether or not the person is drinking beer is on one side of the card and the person's age is on the other side. For two of the people, you can see their age, but you cannot see what they are drinking. For the other two people, you can see what they are drinking, but you cannot see their age. Your task is to decide whether or not this law is being broken in the bar. Circle the card or cards you would definitely need to turn over to decide whether or not the law is being broken. You may select any or all of the cards.

Please indicate your choice or choices by circling either the 'a' (turn) or 'b' (not turn) below each card.

Drinking Beer	16 years	Drinking Beer	19 years

APPENDIX A2.1 Inventory to determine typical/untypical traits (Exp.2 Pilot study)

Occupations Inventory

Below you will find a list of thirty-four different professions/occupations. Next to each one is a list of six character traits or behaviours. Could you please indicate on the scale by each one, how typical the traits/behaviours are for persons in each profession/occupation from 1 to 5 (1 = very untypical, 2 = quite untypical, 3 = neither typical nor untypical, 4 = quite typical, 5 = very typical).

		Very Untypical 1	2	3	4	Very Typical 5
1. Nurses are	**caring**					
	aggressive					
	thoughtful					
	intelligent					
	lazy					
	healthy					
2. Boxers are	**aggressive**					
	quiet					
	talkative					
	healthy					
	demonstrative					
	shy					
3. Firemen are	**brave**					
	intelligent					
	lazy					
	cowardly					
	athletic					
	selfish					
4. Teachers are	**clever**					
	artistic					
	extrovert					
	child hitters					
	boring					
	scientific					
5. Politicians are	**trustworthy**					
	good with children					
	good public speakers					
	analytical					
	fun loving					
	very intuitive					
6. Doctors are	**logical**					
	clever					
	caring					
	intuitive					
	healthy					
	trustworthy					

7. Dentists are	healthy caring artistic talkative happy quiet
8. Writers are	pragmatic intelligent creative disorganised analytical introverted
9. Artists are	loners imaginative logical hard working lazy egocentric
10. Shop assistants are	good with money talkative polite quiet smartly dressed snobby
11. Paramedics are	courageous healthy caring fast thinkers good drivers team workers
12. Musicians are	disorganised creative clever logical introvert egocentric
13. Managers are	organised conforming creative leaders honest overweight
14. I.T. Consultants are	analytical creative intelligent strong wealthy intuitive

15. Telephone operators are	polite talkative caring happy wealthy organised
16. Taxidrivers are	safe drivers drinkers imaginative happy slow drivers helpful
17. Waiters are	polite intelligent caring extrovert organised snobby
18. Nannies are	aggressive domineering caring healthy child lovers creative
19. Librarians are	extrovert organised short-sighted healthy book lovers quiet
20. Bus drivers are	polite aggressive creative pragmatic caring good at maths
21. Soldiers are	brave intelligent power crazy cowardly flirtatious healthy
22. Hairdressers are	conceited creative funny analytical boring intelligent

23. Plumbers are	Socially skilled reliable practical brave wealthy hard working
24. Electricians are	lazy socially skilled problem solvers reliable hard working creative thinkers
25. Engineers are	analytical lazy good at group work creative hard working intelligent
26. Ski instructors are	flirtatious extroverted fun loving reliable boring shy
27. Aerobic instructors are	healthy energetic vain extroverted unhealthy sun-lovers
28. Secretaries are	good organisers gossips helpful common intelligent caring
29. Aeroplane pilots are	intelligent posh brave rational intuitive drinkers
30. Carpenters are	creative artistic religious strong problem solvers analytical

31. Builders are	**bullies** **sexist** **extroverts** **common** **rich** **politically correct**
32. Solicitors are	**snobby** **elitist** **pragmatic** **honest** **down to earth** **impractical**
33. Salespersons are	**common** **talkative** **socially skilled** **money grabbing** **quiet** **generous**
34. Accountants are	**good with numbers** **happy** **boring** **socially skilled** **really interesting** **extroverts**

APPENDIX A2.2 Descriptive statistics for traits related to occupations (highest mean first)

	Valid N	Mean	Min	Max	Std.Dev.
ACCOUNT1	18	4.833333	4	5	0.383482
LIB5	18	4.777778	4	5	0.427793
NURSES1	18	4.722222	4	5	0.460889
ARTIST2	18	4.722222	4	5	0.460889
FIREMEN1	18	4.666667	4	5	0.485071
WRITERS3	18	4.666667	4	5	0.485071
AEROBIC1	18	4.611111	4	5	0.501631
AEROBIC2	18	4.611111	4	5	0.501631
NANNIE5	18	4.611111	3	5	0.607685
DOCTORS2	18	4.555556	4	5	0.51131
MUSIC2	18	4.5	4	5	0.514496
PLUMBER3	18	4.5	4	5	0.514496
POLITIC3	18	4.444444	3	5	0.615699
AEROP1	18	4.444444	3	5	0.615699
BOXERS1	18	4.388889	2	5	0.777544
NANNIE3	18	4.388889	3	5	0.607685
SOLDIER1	18	4.388889	3	5	0.607685
ENGIN1	18	4.388889	3	5	0.607685
PARA6	18	4.333333	3	5	0.594089
SECRET1	18	4.333333	3	5	0.594089
IT1	18	4.277778	3	5	0.669113
PARA3	18	4.277778	3	5	0.574513
LIB2	18	4.277778	3	5	0.574513
SALES2	18	4.277778	4	5	0.460889
MANAGE4	18	4.222222	3	5	0.732084
ENGIN6	18	4.222222	3	5	0.548319
PARA1	18	4.166667	3	5	0.707107
PARA4	18	4.166667	2	5	0.857493
LIB6	18	4.111111	3	5	0.758395
AEROP4	18	4.111111	3	5	0.471405
PARA5	18	4.055556	3	5	0.639137
IT3	18	4.055556	3	5	0.639137
SKI1	18	4.055556	3	5	0.725358
SKI3	18	4.055556	3	5	0.539305
SALES3	18	4.055556	3	5	0.639137
SKI2	18	4	3	5	0.685994
FIREMEN5	18	3.944444	3	5	0.639137
MANAGE1	18	3.944444	2	5	0.802366
AEROBIC4	18	3.944444	3	5	0.539305
BOXERS4	18	3.888889	1	5	0.963382
TEACHERS	18	3.888889	2	5	0.832352
WRITERS2	18	3.888889	3	5	0.471405
WAITER1	18	3.888889	3	5	0.6764
SOLDIER6	18	3.888889	1	5	1.131833
BUILD2	18	3.888889	3	5	0.758395
SOLIC2	18	3.888889	1	5	0.963382
DOCTORS1	18	3.888889	2	5	0.832352
ELEC3	18	3.888889	2	5	0.582983
IT5	18	3.833333	2	5	1.043185
ARTIST1	18	3.833333	3	5	0.707107
TELE1	18	3.833333	3	5	0.785905
CARP1	18	3.833333	2	5	0.923548
HAIR2	18	3.722222	2	5	0.751904
BUILD4	18	3.722222	3	5	0.669113

DOCTORS6	18	3.722222	2	5	0.826442
SHOP2	18	3.722222	2	5	0.826442
MUSIC6	18	3.666667	3	5	0.594089
TELE2	18	3.666667	3	5	0.685994
ENGIN5	18	3.666667	3	4	0.485071
CARP4	18	3.666667	3	4	0.485071
SALES4	18	3.666667	1	5	1.084652
ARTIST6	18	3.666667	1	5	0.970143
SOLIC1	18	3.666667	1	5	0.907485
ACCOUNT3	18	3.611111	1	5	1.036901
NURSES3	18	3.611111	3	5	0.697802
AEROBIC3	18	3.611111	3	5	0.607685
SECRET2	18	3.611111	3	5	0.697802
CARP5	18	3.611111	3	5	0.607685
BUILD3	18	3.555556	3	5	0.615699
SHOP5	18	3.5	3	5	0.618347
SOLDIER5	18	3.5	2	5	0.707107
AEROBIC6	18	3.5	3	5	0.618347
CARP2	18	3.5	2	5	0.857493
NURSES4	18	3.444444	3	4	0.51131
WRITERS4	18	3.444444	1	5	0.921777
MUSIC3	18	3.444444	3	4	0.51131
BOXERS5	18	3.388889	1	5	1.036901
TEACHERS	18	3.388889	2	5	0.697802
POLITIC4	18	3.388889	2	4	0.777544
DENTISTS	18	3.388889	2	4	0.607685
MANAGE2	18	3.388889	2	5	0.697802
IT6	18	3.388889	2	4	0.607685
WAITER5	18	3.388889	2	4	0.697802
ENGIN3	18	3.388889	2	5	0.697802
SECRET3	18	3.388889	1	4	0.777544
AEROP3	18	3.388889	2	5	0.777544
AEROP5	18	3.388889	2	4	0.777544
CARP6	18	3.388889	2	4	0.607685
DOCTORS3	18	3.388889	1	5	0.978528
SHOP1	18	3.333333	2	5	0.685994
SOLDIER3	18	3.333333	2	5	0.766965
ELEC5	18	3.333333	2	4	0.685994
PLUMBER6	18	3.333333	1	5	0.970143
AEROP2	18	3.333333	1	5	0.907485
SHOP3	18	3.277778	1	5	0.95828
TELE6	18	3.277778	2	4	0.574513
WAITER4	18	3.277778	1	4	0.751904
TEACHERS	18	3.277778	3	4	0.460889
ENGIN4	18	3.277778	1	5	1.017815
SOLIC3	18	3.277778	1	4	0.826442
WRITERS6	18	3.222222	2	4	0.808452
FIREMEN2	18	3.166667	2	4	0.707107
TEACHERS	18	3.166667	2	5	0.618347
DENTISTS	18	3.166667	2	5	0.857493
SHOP6	18	3.166667	2	4	0.707107
PARA2	18	3.166667	2	4	0.514496
BUS1	18	3.166667	1	5	0.985184
SALES1	18	3.166667	2	4	0.514496
DOCTORS4	18	3.166667	1	4	0.785905
IT2	18	3.166667	1	5	0.923548
NANNIE6	18	3.111111	1	5	0.900254
PLUMBER5	18	3.111111	2	5	0.900254
TEACHERS	18	3.111111	2	4	0.6764
NANNIE4	18	3.111111	2	4	0.471405

SECRET5	18	3.111111	1	4	0.900254
BUILD1	18	3.111111	1	5	0.963382
DOCTORS5	18	3.055556	2	4	0.639137
DENTISTS	18	3.055556	2	5	0.639137
MUSIC1	18	3.055556	2	4	0.639137
MANAGE6	18	3.055556	1	5	0.725358
BOXERS3	18	3	2	4	0.594089
WRITERS5	18	3	1	5	0.907485
ARTIST5	18	3	2	4	0.685994
TAXI4	18	3	1	4	1.084652
TAXI6	18	3	1	5	1.236694
WAITER6	18	3	1	4	0.840168
NANNIE2	18	3	1	5	1.028992
LIB3	18	3	1	5	0.766965
BUS4	18	3	1	4	0.840168
HAIR3	18	3	1	4	0.766965
HAIR5	18	3	2	5	0.840168
SECRET6	18	3	1	4	0.685994
BUILD5	18	3	1	5	0.907485
NURSES6	18	2.944444	1	5	0.872604
WRITERS1	18	2.944444	1	4	0.937595
HAIR1	18	2.944444	1	4	0.725358
TELE4	18	2.944444	1	5	0.872604
BUS2	18	2.944444	2	4	0.802366
DENTISTS	18	2.888889	1	4	0.6764
ARTIST4	18	2.888889	2	4	0.832352
BUS5	18	2.888889	1	4	0.900254
DENTISTS	18	2.888889	2	4	0.6764
MANAGE5	18	2.888889	1	4	0.758395
SECRET4	18	2.888889	2	5	0.758395
TAXI1	18	2.833333	1	5	1.294786
TAXI2	18	2.833333	1	4	0.857493
MUSIC4	18	2.777778	1	4	0.808452
LIB4	18	2.777778	1	3	0.548319
ELEC2	18	2.777778	1	4	0.646762
SOLIC6	18	2.777778	1	4	0.732084
WAITER2	18	2.722222	1	3	0.574513
BUS6	18	2.722222	1	4	0.751904
ACCOUNT4	18	2.722222	2	4	0.669113
MUSIC5	18	2.666667	2	4	0.594089
MANAGE3	18	2.666667	2	4	0.594089
SKI4	18	2.666667	2	4	0.594089
WAITER3	18	2.666667	1	4	0.766965
POLITIC6	18	2.611111	1	4	0.916444
ENGIN2	18	2.611111	2	4	0.607685
ACCOUNT2	18	2.611111	2	3	0.501631
SHOP4	18	2.611111	1	3	0.607685
PLUMBER1	18	2.611111	1	4	0.697802
ELEC1	18	2.611111	1	4	0.697802
POLITIC2	18	2.555556	1	4	0.704792
ELEC4	18	2.555556	1	4	0.783823
CARP3	18	2.555556	1	3	0.783823
DENTISTS	18	2.5	1	4	0.857493
TELE3	18	2.5	1	3	0.618347
BUS3	18	2.5	1	3	0.618347
SOLDIER2	18	2.5	1	4	0.785905
ELEC6	18	2.5	1	4	0.857493
POLITIC5	18	2.444444	1	3	0.615699
TAXI3	18	2.444444	1	4	0.855585
PLUMBER4	18	2.444444	1	3	0.783823

ACCOUNT6	18	2.444444	1	3	0.615699
IT4	18	2.388889	1	4	0.849837
AEROP6	18	2.388889	1	4	0.849837
SKI5	18	2.388889	2	3	0.501631
BOXERS2	18	2.333333	1	4	0.970143
SOLIC4	18	2.333333	1	4	1.028992
SALES6	18	2.333333	1	3	0.840168
ARTIST3	18	2.277778	1	3	0.669113
HAIR6	18	2.222222	1	3	0.878204
BOXERS6	18	2.166667	1	4	0.923548
FIREMEN3	18	2.166667	1	4	0.857493
FIREMEN6	18	2.166667	1	3	0.707107
TELE5	18	2.111111	1	3	0.758395
SOLIC5	18	2.111111	1	3	0.758395
NURSES5	18	2.055556	1	5	1.258955
SOLDIER4	18	2.055556	1	3	0.639137
HAIR4	18	2.055556	1	3	0.802366
SALES5	18	2	1	4	0.766965
ACCOUNT5	18	2	1	3	0.766965
PLUMBER2	18	1.944444	1	3	0.802366
LIB1	18	1.888889	1	3	0.832352
POLITIC1	18	1.888889	1	4	0.963382
TAXI5	18	1.833333	1	3	0.785905
SKI6	18	1.833333	1	3	0.707107
NANNIE1	18	1.777778	1	4	0.942809
NURSES2	18	1.722222	1	4	0.95828
BUILD6	18	1.722222	1	3	0.751904
FIREMEN4	18	1.611111	1	5	0.978528
AEROBIC5	18	1.611111	1	3	0.697802
TEACHERS	18	1.5	1	3	0.707107

APPENDIX A2.3 Descriptive statistics of typical/untypical/neutral traits

Table 1. Means and Standard deviations of Typical character traits for occupations to be used in Experiment 2 (N=18).

Occupation/ trait	Mean	Minimum	Maximum	Std.Deviation
Accountant-Good with Numeracy	4.83	4	5	0.38
Librarian-Book lovers	4.78	4	5	0.43
Nurses-Caring	4.72	4	5	0.46
Artists-Imaginative	4.72	4	5	0.46
Firemen-Brave	4.67	4	5	0.48
Writers-Creative	4.67	4	5	0.48

Table 2. Means and Standard Deviations of Untypical character traits for occupations to be used in Experiment 2 (N=18).

Occupation/ trait	Mean	Minimum	Maximum	Std.Deviation
Teachers-Hit children	1.5	1	3	0.71
Aerobic Instructors-Unhealthy	1.61	1	3	0.70
Firemen-Cowardly	1.61	1	5	0.98
Builders-Politically Correct	1.72	1	3	0.75
Nurses-Aggressive	1.72	1	4	0.96
Ski Instructors-Shy	1.83	1	3	0.71

Table 3. Means and Standard Deviations of Neutral character traits for occupations to be used in Experiment 2 (N=18).

Occupation/ trait	Mean	Minimum	Maximum	Std.Deviation
Boxers-Talkative	3	2	4	0.59
Artists-Lazy	3	2	4	0.68
Taxi-Drivers-Helpful	3	1	5	1.24
Nannies-Domineering	3	1	5	1.03
Librarians-Shortsighted	3	1	5	0.77
Secretaries-Caring	3	1	4	0.69

Law of Large Numbers Written Explanation

We are very interested in studying how people go about explaining and predicting events under conditions of very limited information about the events. It seems to us to be important to study how people explain and predict under these conditions because they occur very frequently in the real world. Indeed, we often have to make important decisions based on such explanations and predictions, either because there is too little time to get additional information or because it is simply unavailable.

Experts that study human inference have found that some common-sense principles are helpful in explaining and predicting events, especially under conditions of limited information. One such principle of probability that is particularly helpful is called the *Law of Large Numbers.* In this study, we will teach you the Law of Large Numbers by introducing you to the probabilistic terms and ideas associated with this principle, and then provide examples to illustrate how the Law of Large Numbers can be used to explain and predict events in the real world.

Imagine an urn that is filled with gumballs. Let's say that the urn contains a very large number of gumballs - thousands, millions, or larger. The gumballs in this urn are known collectively as the population.

Let's say that there are two types of gumballs in the urn - red gumballs and white gumballs. When we do this, we can now say that the population has two categories or groups, namely - red gumballs and white gumballs.

Now let's say that in this population of red and white gumballs there are 70% white gumballs and 30% red gumballs. If that is the case, then we know more than that the population has two groups (red and white): we now know the proportion of the gumballs in the white group (70%) and the proportion of the gumballs in the red group (30%). This is known as the population distribution (other examples of distributions are 60% red and 40% white, or 85% white and 15% red, etc., but in every distribution, the sum of the proportions must be 100%.

Let's summarise what we've covered so far:
- A population is the entire set of objects we are interested in (all of the gumballs in the urn).
- Groups refer to the types of objects in the population (red and white).
- Distribution refers to the proportion of objects in each group (70% white and 30% red in this example.

One of the major goals of statistics is to find out something about a population. More specifically, we want to find out what the population distribution is. One way that we might do this would be to actually examine all of the objects in the population and count up the number of objects in each group. In our example, we would empty the entire urn and count the number of red gumballs and the number of white gumballs. Using this method, we could find out exactly what the population distribution of red and white gumballs was. But, there is a very serious problem with counting all of the objects in the population: populations, in general, are very large. If we were to count all of the objects in our gumball population, it would take more time and effort than would be practical (imagine counting a million gumballs!)

OK - so counting the entire population is impractical. What do we do instead to find out what the population distribution is?

What we do instead is to take a sample of the population. A sample is a subset of the population. We can take a sample of any size - if we pick 5 gumballs, we say that the sample size is 5; if we take 60 gumballs for our sample, we say that the sample size is 60, and so on.

When we take a sample from the population, we will get a sample distribution. The sample distribution is the proportion of objects in each group for the sample, just as the population distribution is the proportion of objects in each group for the population. For example, if we take a sample of 10 gumballs, we might get 6 reds and 4 whites. In this case, our sample distribution would be 60% red and 40% white. We also might have happened to get 9 whites and 1 red, in which case the sample distribution would be 90% white and 10% red. The important point here is that samples are estimates of populations. Since it is often impractical or sometimes impossible to examine the entire population, we instead have to draw sample to estimate what the population is like.

Some samples will have sample distributions that are closer to the population distribution than others. For instance, in our gumball example, a sample of 9 reds and 1 white would be a very poor estimate of the population, while a sample of 8 whites and 2 reds would be a pretty good estimate of the population. The critical question is: *What determines how likely it is that samples will give good estimates of the population*? The answer is simple: if the samples are chosen haphazardly, or randomly (by, for example, mixing the urn and reaching into the urn blindfolded and scooping out the needed number of gumballs OR by mixing the contents of a gumball machine and letting the gumballs out one by one), then there is only one factor - *sample size.*

This brings us to the Law of Large Numbers: as the size of a random sample increases, the sample distribution is more and more likely to get closer and closer to the population distribution. In other words, the larger the sample, the better it is as an estimate of the population.

Verbal presentation of the Law of Large Numbers

Now that you have all read the written explanation of the Law of Large Numbers, I thought it would be nice to demonstrate, before your eyes, that the Law of Large Numbers really does work.

So I have here (pick up jar) a genuine jar, filled with genuine red and white beads. And as in the written explanation, there happen to be 70% white and 30% red beads in this jar.

A major purpose of statistics is to find out about a population from a sample of that population. Suppose it is your job to find out what proportion of the beads in this jar are white and what proportion of the beads are red. You could dump out all of the beads and count all of them, but that would take quite a long time and wouldn't be worth the effort. For the sake of demonstration, this jar isn't very large, but if we had a very large jar filled with millions of beads, it's easy to see how time-consuming and impractical it would be to count the entire population of beads in the jar.

What you would probably do instead to find out what the composition of the jar was like would be to take a sample from the jar because the sample you chose would tell you something about the population; that is, the sample would be an estimate of the population.

According to the Law of Large Numbers, when you choose your sample randomly like this (reach into jar without looking, mix them up, and draw out a handful), the larger the sample, the better the sample is in estimating the population. To repeat - the larger the sample is that you draw, the better that sample is in estimating the population.

Well, what I'm going to do now is to demonstrate the Law of Large Numbers (reveal blackboard/whiteboard with summary chart). I will pick samples of size 1, 4, and 25 (gestures to the three sections of the board as you say the numbers) to show that as the sample size increases, the sample becomes a better estimate of the population.

For each sample that I draw, I will write down on the board the number of whites in the sample, the number of reds in the sample, the percent white in the sample, and the deviation or difference between the sample distribution and the population distribution (as you are saying the various categories, point to them on the board).

Now recall that the population distribution for this jar is 70% white, 30% red. Therefore, for example, if the sample I draw happens to be 85% white, the deviation of that sample will be 85 minus 70 or 15%, and I'll enter that number here (point to the deviation column).

I will draw a draw a few samples of size 1, 4, and 25. After I'm done with drawing samples of each size I will calculate the average deviation of the samples from the population (point to all three 'Average Deviation' boxes). The Law of Large Numbers states that as the sample size increases, the sample becomes a better estimate of the population. In other words the average deviation of the sample from the population will decrease as the sample size increases. So this number (point to 'Average Deviation' box) should go down as sample size (point to top of chart: 'Sample Size = ' goes up.

So first I will draw samples of size 1. (take scooper out of jar). I will mix up the beads like this (mix beads while talking) and use this scooper to pick my sample (put scooper in, get some beads).

OK, the first gumball to come out of the scoop will be my sample. In this sample, I have 0 whites and 1 red (for example). (Go to the board and verbalise as you're writing down the results of the sample) That means that percent white in the sample was 0%. 0 minus 70 equals 70% deviation.

(go back to jar. Put the sample back and repeat the procedure three times)

(after you've finished, compute the average deviation) So the average deviation for samples of size 1 is _ %.

Now I'll pick a few samples of size 4. (Follow the same procedure as above - mix contents of jar, scoop out four beads, summarise, put sample back in the jar and repeat three more times before computing the average deviation.) The average deviation for samples of size 4 is _ %.

(do the same with samples of size 25. Three should be enough. With samples of size 25 you can't hold 25 beads in one hand, you will have to shake a few beads into your hand at a time. Use two glass bowls and when you shake a few out, separate the whites and reds and put them into different bowls. Keep track of how many you've already drawn so that you end up with 25.)

(Example) With this sample of 25 I have 18 whites and 7 reds (write results on board). That is 72% white and the deviation is 2%.

(if you have small deviations, you can quit at 3 samples of 25)

(compute average deviation for samples of size 25)

(it is possible, but not likely, that samples of size 25 will give larger deviations than samples of size 4, especially if size 4 gave you all 3-1 sample splits (average deviation = 5%)
Two ways to guard against this;

1. If you've drawn 4 samples of size 4 and they're all 3-1, pick more samples until you have some other sample distribution.
2. If you've drawn 3 samples of size 25 and you think your average deviation might be too close to the average deviation for samples of size 4, draw one or two more.

Summary

The Law of Large Numbers states that as the size of your sample increases, the sample becomes a better estimate of the population. This is shown here: as the samples increased in size from 1 to 4 to 25 (gesture to the top of the chart).

I'd like to tell you something else about the Law of Large Numbers. That is, with small samples, sometimes you can't even correctly answer the simplest questions about the population.

For example, suppose you were asked to say whether there were more white beads in the jar or more red beads. If you happened to draw this sample (point to a very bad sample of size 1: 0 white, 1 red. If there isn't one, then go to the samples of size 4 and point to a 1 white, 4 red or 2 white, 2 red) you would say ' Well, from my sample, I think there are more reds than whites' or 'I can't tell at all' and you would be wrong. But look at the larger sample of size 25: you can always correctly answer at least the most basic question - 'Are there more whites or more reds?' With smaller samples, that is not always possible.

So I've demonstrated the Law of Large Numbers - as the sample increases in size, the sample becomes a better estimate of the population.

Examples training

One reason that the Law of Large Numbers is important to learn is that it applies not only to urns and gumballs. The basic principles involved in the law of large numbers apply whenever you make a generalisation or an inference from observing a sample of objects, actions or behaviours.

To give you an idea of how broad the law of large numbers is, we have, in this packet, presented three situations in which the law of large numbers applies. Each situation is analysed in terms of the law of large numbers.

Please read each problem and consider it for a while before turning the page to read the Law of Large Numbers answer.

Structure 1: Generalising from a small sample

A major London law firm had a history of hiring only graduates of large, prestigious law schools. One of the senior partners decided to try hiring some graduates of smaller, less prestigious law schools. Two such people were hired. Their grades and general record were similar to those of people from the prestigious schools hired by the firm. Although their manners and 'style' were not as polished and sophisticated as those of the predominantly Cambridge junior members of the firm, their objective performance was excellent. At the end of 3 years, both of them were well above average in the number of cases won and in the volume of law business handled. The senior partner who had hired them argued to colleagues in the firm that, "This experience indicates that graduates of less prestigious schools are at least as ambitious and talented as graduates of the major law schools. The chief difference between the two types of graduates is in their social class background, not in their legal ability, which is what counts."

Comment on the thinking that went into this senior partner's conclusion. Is the argument basically sound? Does it have weaknesses? (Disregard your own initial opinion, if you had one about graduates of nonprestigious law schools, and concentrate on the thinking that the senior partner used).

Analysis

The senior partner is trying to draw a conclusion about a certain population. We can think of the members of this population as newly graduated solicitors, from nonprestigious law schools, who otherwise meet the law firm's hiring standards. If we divide the members of this population into two groups, 'excellent' and 'mediocre or worse' we can think of the population distribution as the percentage in each group. The senior partner has concluded that the percentage in the 'excellent' group is very high, or anyway, just as high as in another population, involving graduates of prestigious law schools. This conclusion was based on observing a sample of size = 2, in which the sample distribution was 100% 'excellent', 0% 'mediocre or worse'.

Apart from any other considerations, however, the sample distribution for size 2 is apt to be quite different from the population distribution: the latter could be only 60 or 50% or even perhaps as low as 40% 'excellent', and a 2-0 sample split would not be so unusual; just as one would not be at all amazed to draw two out of two red gumballs from an urn with only 40% reds. So the senior partner's attitude is quite unwarranted: a larger sample is needed.

Structure 2: Regression

Susan is the artistic director for a ballet company. One of her jobs is auditioning and selecting new members of the company. She says the following of her experience: "Every year we hire 10-20 young people on a 1-year contract on the basis of their performance at the audition. Usually we're extremely excited about the potential of 2 or 3 of these young people - a young woman who does a brilliant series of turns or a young man who does several leaps that make you hold your breath. Unfortunately, most of these young people turn out to be only somewhat better than the rest. I believe many of these extraordinarily talented young people are frightened of success. They get into the company and see the tremendous effort and anxiety involved in becoming a star and they get cold feet. They'd rather lead a less demanding life as an ordinary member of the corps de ballet."

Comment on Susan's reasoning. Why do you suppose that Susan usually has to revise downward her opinion of dancers that she initially thought were brilliant?

Analysis

We can analyse this problem using the laws of large numbers by thinking of each ballet dancer as possessing a <u>population</u> of ballet movements. Susan is interested in excellence, so we can divide the members of each population into two categories: 'brilliant movements' and 'nonbrilliant, or other movements'. We can think of the <u>population distribution</u> as the percentage or proportion in each category. For many dancers, the population distribution is actually 0% brilliant and 100% other; these dancers simply lack the talent to perform brilliant movement. For many other dancers, there is a small or moderate percentage of 'brilliant movement' gumballs in their urn. A true ballet star would therefore have a population distribution with a greater percentage of 'brilliant' movements than an ordinary member of the corps de ballet.

By conducting auditions, Susan is observing <u>samples</u> of each dancer's population distribution. An audition, however, is a very small sample of a dancer's movements. We know from the law of large numbers that small samples are very unreliable estimates of the population. When a dancer performs some brilliant moves during an audition, it is often because the dancer has happened to draw a couple of the 'lucky gumballs' that day: it does not prove that the population distribution for that dancer consists of a large percentage of 'brilliant movements'. It is reasonable to think that there are really very few dancers that have population distributions with a large percentage of brilliant movements; and so when Susan sees a dancer performing brilliantly at audition, the chances are it is just a lucky draw from a dancer who is capable of performing some, but not necessarily a great number of 'brilliant movements'. Therefore, when Susan hires such dancers and evaluates them after seeing a much larger sample of their movements, it is not surprising that she finds that many of these dancers that were brilliant at audition turn out to be only somewhat better than the rest.

Structure 3: Large sample vs. theory without supporting data

Kevin, a graduate student in sociology, decided to do a research project on 'factors affecting performance of 1st division football players', in which he gathered a great amount of demographic data on birthplace, education, marital status, etc. to see if any demographic factors were related to the performance of 1st division football players (e.g. goal scoring average, penalties). Kevin was unable to use data for all the 1st division teams because information for some of the players was unavailable, but he was able to obtain data for some 200 players in the 1st division.

One finding that interested Kevin concerned the 110 married players. About 68% of these players improved their performance after getting married, while the remainder had equal or poorer performance. He concluded that marriage is beneficial to a football player's performance. At a social hour sponsored by the Football Association, he happened to mention his finding to a staff member of the office. The staff member listened to Kevin's results and then said, 'Your study is interesting but I don't believe it. I'm sure that football performance is worse after a marriage because the football player suddenly has to take on enormous responsibilities: taking care of his spouse and children. Plus the factor of being stressed by having to be on the road so much of the time and therefore away from the family. The player will no longer be able to devote as much time to football as before he was married. Because of this he will lose that competitive quality that is necessary for good performance in football.

What do you think of the staff member's argument? Is it a sound one or not? Explain your reasoning.

Analysis

Kevin is trying to find out how performance in 1st division football is affected by being married. To do this, he obtained data for 200 players in the premier league and discovered that out of the 110 that had gotten married, 68% had improved performance after the wedding (and 32% had equal or poorer performance). According to the law of large numbers, which states that the larger the sample, the better it is in estimating the population, there is substantial evidence that marriage is beneficial to football players' performance. Recall that in the gumball demonstration, samples of size 25 were very good estimates of the population: these samples did not differ much from population. Extending the argument, samples of size 110 are extremely accurate estimates of the population. Thus, it can be concluded that, in general, marriage is beneficial to football players' performance.

What about the staff member's theory that football performance is worse after a marriage because the player assumes enormous responsibilities and will no longer be able to devote as much time to football as before? Although this argument may have some intuitive appeal, it should be discounted because it is not supported by any data and is, in fact contradicted by Kevin's large sample of 110 players.

APPENDIX A2.5 Law of Large Numbers reasoning problems (Experiment 2 and 3)

Instructions

On the pages that follow, there are a number of arguments that we would like you to consider. As you will see, they represent a wide range of real-life situations. We would like you to think carefully about each argument, and then write down answers that are sensible to you.

Following each argument could you please rate the strength of the conclusion (1=extremely weak; 9=extremely strong) and how persuaded you are by the argument (1=extremely unconvinced; 9=extremely convinced). Space is provided under each argument for you to write your explanations. In many of the arguments, you may find that the Law of Large Numbers is helpful.

If you have any questions, please ask the experimenter.

Neutral Problems

Structure 1:

John and Louise are having an argument about what occupation John should take up when he leaves school. Louise says, 'Listen if you want to earn loads of money you should go into the building trade! My Dad and my Uncle are both builders and they're rolling in it! They've both been doing it for quite a while now and look at them. New big houses, two holidays a year, flash cars......! If your goal is to make lots of money then become a builder'.

Structure 2:

Fiona and Simon are arguing over whether Librarians have more eyesight problems than other people.

Fiona's argument is, "Librarians don't have more eye-sight problems than people from any other profession. I visit the library at least once a week and there are no more people wearing glasses than anywhere else! Do you know Mrs Travis? She's worked in the local library for years and she doesn't wear glasses".

Simon's argument is, "Well I was reading this article the other day that stated that people who had to read a lot as part of their job were very likely to become short-sighted. As part of the article a large study on thousands on people from a variety of occupations was conducted which showed specifically that Librarians as well as some other professions, were more likely to be shortsighted.

Structure 3:

Suzi is looking for a Nanny to look after her one-year-old son while she is at work. She has shortlisted the applicants down to two and they are to each have an interview followed by an hour's play with her son so that Suzi can observe how they interact with him.

At the interviews, Suzi made notes on how each prospective Nanny interacted with her son. Nanny 1 had been very jolly and had helped him decide what to wear that day and then organised what games to play. Nanny 2 had fed Suzi's son his dinner and encouraged him to feed himself his pudding.

Suzi concluded at the end of the interviews that she wasn't sure about returning to work after all, or maybe a creche would be better as Nannies have very domineering personalities.

Untypical Problems

Structure 1:

An editorial in a local newspaper recently criticised the occupation of being an Aerobics instructor. The journalist's argument was:

I've got a friend who's an aerobic instructor and I wouldn't want anyone I know to copy her lifestyle. She is so unhealthy. She drinks and smokes and actually never takes any real exercise herself, she just tells others how to do so! I know her flatmates and they say she never stops eating as well, not healthy food either, fry ups and chocolate are normal. My conclusion? I don't think there's a more unhealthy group of people than Aerobics instructors!

Structure 2:

James and Kim are having an argument over whether Teachers are more likely to hit children than other people. James claims that the children of Teachers are more unhappy than other children and frequently have unexplainable bruises, but Kim believes that Teacher's children are more happy and well adjusted than other children.

Kim:
I've been working as a developmental psychologist for fifteen years and handle child abuse cases all the time. I've seen hundreds of children who have been beaten by their mother or father who have worked in all kinds of professions, but very few indeed have been children of Teachers. I've seen enough to know that a child is very likely to be happy and secure if at least one of their parents is a Teacher.

James:
You're wrong and you know it! Two people who live in our road are teachers and they regularly beat their children. They are always covered in bruises. In fact I know some other teachers who hit their children much more than they should, it'd disgusting. You see, I've seen it for myself. If you'd seen what I've seen you would agree that people who are Teachers are more likely to hit children.

Structure 3:

David did his thesis in sociology on the subject of bravery in the face of danger in different occupational groups. In his study, 3 engineers, 3 firemen and 3 ambulance men participated. Below is a summary of David's study.

When they first came into the laboratory, each of the participants had to complete a questionnaire asking them what risks they would take to help others in a variety of different situations where their life could be put in danger if they tried to save them. Then they were asked to complete a range of cognitive tasks during which alarms would ring without warning. At the same time the participants' heart rates were being monitored to observe physiological reactions. The results were as follows:

Overall, the 3 engineers had the highest score for the risk-taking questionnaire and the lowest mean heart rate when the alarms sounded. The firemen on the other hand had the lowest mean score for risk-taking and the highest mean heart rate score when compared to the other two professions. David concluded in his thesis that firemen are much more cowardly than either engineers or ambulance men.

Typical Problems

Structure 1:

Paul is discussing relationships with his sister Kate. His girlfriend at the moment is a Nurse and he likes her a lot. She is great fun to be with and Paul thinks she is very caring.

Paul: I've never been so happy in my life! She is never snappy or abrupt, she always asks how I am and really listens to what I have to say. The other night she new I was going to be home late so she brought round a meal all ready cooked so I just had to heat it up. If you could find yourself a male nurse I would because I have never known such a caring person to have a relationship with! I really recommend it. Nurses are really caring.

Structure 2:

Dr Avon is a psychologist who has spent several years determining whether people in certain occupations are more imaginative than people in other occupations. In a recent study, he compared approximately 500 artists to approximately 500 sales persons. In the report he published, Dr Avon found that:

"It is clear from these results that artists are much less imaginative than sales persons. Artists have very fixed ideals, are narrow minded and have very strict boundaries within which they allow themselves to work. On the other hand, salespersons are very imaginative, using different methods and strategies to perform in their role."

In response to his paper, Dr Bing, also a respected psychologist, wrote a paper in which he attempted to discredit Dr Avon's argument. Dr Bing wrote:

"I have difficulty believing Dr Avon's study for two reasons. First, I share my house with an artist who has a very vivid imagination. Second, talking about the 'average' behaviour of artists doesn't tell me anything about how any one artist is. On this basis, I feel we can effectively dismiss Avon's study."

Structure 3:

Laura works for a London testing company. Her company has been hired to measure the Cognitive Abilities of Ministry of Defence employees. Laura herself has been asked to administer the tests to 10 people who are MOD employees. Two people were secretaries, two were doctors, two were accountants, two were dentists and two were architects. A significant section of the test was on Numeracy and she used the best, one that is the most respected across the country. After each person was tested, she put the findings in a table:

Person	Numeracy Score
Secretaries	56 and 63
Doctors	69 and 72
Accountants	89 and 95
Dentists	65 and 70
Architects	75 and 79

Because a person with average cognitive ability should have a Numeracy score of around 80 and because below that is considered low, Laura concluded that Accountants are much better working with numbers than the other groups of people she tested.

APPENDIX A2.6 Experiment Evaluation Problems (Experiments 2 and 3)

Instructions

On the following pages you will be presented with several summaries of actual psychological and sociological research involving participants from several dozen occupations.

Following each problem could you please rate the strength of the researcher's conclusion (1=extremely weak; 9=extremely strong) and the validity of the experiment i.e. does the experiment measure what the researcher suggests it measures (1=extremely invalid; 9=extremely valid). Space is also provided under each problem for you to write an explanation of why you believe the researcher's conclusion to be weak or strong and the experiment to be valid or invalid.

If you have any questions please ask the experimenter.

Neutral Problems

Structure 1:

Professor Y. wanted to find out whether individuals who worked in any one occupation/profession were more or less talkative and had different social skills than individuals who worked in another. She randomly selected 200 dentists, boxers, door bouncers and actors from all over the United Kingdom and invited them to a conference and dinner together. The Boxers couldn't make it on the same day therefore their conference/dinner was held two weeks later. At the conferences and dinners, individuals were watched by trained researchers to observe their behaviours. For example, number of times they initiated conversation, length of utterances and number of utterances. On collation of all the data, Professor Y. found that the boxers showed the most talkative behaviours, closely followed by dentists, then bouncers and then actors. In reporting her findings, Professor Y. concluded that boxers are much more talkative than dentists, bouncers and actors.

Structure 2:

Dr. S. and Dr. A. designed a study to ascertain whether there were differences in people's activity levels dependent on their occupation. They were particularly interested in people's health and fitness levels. Three hundred nurses, writers and teachers were randomly selected from throughout Devon and Cornwall to participate in the study. Three hundred artists were also recruited from a local country retreat. All participants were required to fill in an attitudes to health and fitness questionnaire followed by an hour of different activities such as swimming, running and cycling whilst having pulse rate and oxygen levels in the blood monitored. Dr. S. and Dr. A. found that the artists had the most negative attitudes towards all aspects of health and fitness and the worst physiological reactions to exercise when compared to the other groups. They concluded in their report that their findings clearly demonstrated that artists are lazy.

Structure 3:

Dr. Wendy had a very strong interest in emotions and more specifically emotional behaviours. She wanted to find out whether people who worked in certain occupations were more able to express their emotions more than others were. Dentists, Teachers and Secretaries were included in the study, 150 of each. Each participant in turn was asked to watch a 15-minute film involving scenes of a disturbing nature, people being hurt, animals in danger and an earthquake. Reactions to the film were measured two ways, through a one-way mirror as they were watching the film and by responses on a questionnaire immediately after the film. It became evident that the majority of the secretaries needed to leave the room at least once during the film, compared to the Dentists and Teachers who didn't feel the need to leave. Also, in the questionnaire, the secretaries reported higher stress scores overall after watching the film. Dr. Wendy concluded from these findings that Secretaries are more caring compared to Dentists and Teachers.

Untypical

Stucture 1:

Dr. H. is a researcher who is interested in finding out how capable people in different occupations are of holding pleasant conversations with strangers. Dr. H. is especially interested in finding out how builders, teachers, solicitors and shop assistants are in dealing with strangers because people in these professions must deal with strangers on a daily basis. To observe their conversations, Dr. H. watched a large number of builders having a conversation with Jim, one of Dr. H's laboratory assistants. Next he watched the teachers, solicitors and shop assistants having a conversation with Amy, another one of his research assistants. Dr. H. noted that teachers, solicitors and shop assistants overall made more inappropriate hand gestures, more body contact with the other person and more wandering eye movements than the builders. He concluded as part of his journal article that builders were more politically correct than any of the other participants from other occupations.

Structure 2:

Two well-known researchers have conducted several studies on the relationship between one's occupation and stress-related behaviours. The occupations of being a teacher and a nurse were compared in one study. In this study, 300 teachers were contacted at local schools and agreed to be in the study. At the time, nurses were more difficult to recruit. Thus, the researchers recruited 300 nurses from a seminar on self-defence that was taking place in a nearby city. Next, each participant was brought into the lab. where they were asked to complete a battery of tests including cognitive ability and stress tests. Participants were then asked to complete a circuit training course whilst blindfolded under the direction of another participant from the same occupation. The two researchers noted that while there wasn't a significant difference between the two groups on the measures of cognitive ability and stress tests, whilst completing circuits, nurses demonstrated many clearly aggressive behaviours. The researchers concluded that nurses are more aggressive.

Structure 3:

Dr. Bill is a psychologist interested in whether people who suffer from shyness are more likely to work in certain occupations. To conduct his research, he included in his study teachers, salespersons, and ski-instructors. 1000 subjects participated in each group. To measure shyness, Dr. Bill attached a microphone to each participant's belt, which the participant wore for 24 hours. The microphone recorded the number of utterances made by the participant in the 24-hour period. At the end of his study, Dr. Bill found that the average teacher spoke 1389 times a day, salespersons spoke a staggering 1520 whilst ski-instructors spoke on average only 250 times a day. Therefore, ski-instructors spoke about 6 times less than people did from the other occupations. Based on this, Dr. Bill concluded that ski-instructors are much shyer than people from other occupations are.

Typical Problems

Structure 1:

A famous sociologist conducted a study looking at personality factors related to people's chosen occupations. He was particularly interested in aspects of some people's personalities that may be more apparent in the workplace for some professions more than others. Large groups of librarians, teachers and nurses were recruited to take part in the study. Each occupational group was asked to attend the lab for an afternoon of testing on different days. Once at the lab, the participants were split into smaller groups of eight. Experienced research assistants who introduced the different topics for discussion led videotaped focus groups. The groups consisting of teachers and nurses were led into discussion about normal working days, recreational activities and stress. The librarians were led into discussion about their normal working day and leisure time also. The researchers noted that the librarians spent a great deal of their time at work and at home with books. The sociologist wrote in his report that librarians love books more than any of the other occupations tested.

Structure 2:

Professor A. recently ran a large-scale study looking at thinking styles in individuals from different occupations. For instance, analytical, pragmatic, deep and shallow thinking styles have all been explored in previous research. With the aid of trained researchers, one hundred and seventy-five writers who were selected from a writing course in the Midlands, and one hundred and seventy-five teachers who were selected from ten large inner city schools in London, were observed for a period of six months going about their work. All participants were also required to keep a diary for the time period of the study, in which they had to write at least one entry per day, describing thoughts that they had had. At the end of the six months the results were collated and generally it seemed to be that writers wrote 50% more entries in their diary than the teachers did and the content was much higher in quality. Practices that the writers had in their daily routine at work also led Professor A. to conclude that writers are very creative in their style of thinking.

Structure 3:

A leading psychologist from a University Hospital conducted a study looking at bravery in the workplace. For the research she compared teachers, nurses, secretaries and firemen. 250 participants were recruited from each occupation and for a fortnight they carried a Dictaphone and a diary and whenever possible they had to describe what they were thinking about. At the end of the study period the results were collated and it was found that overall teachers had an average of 34 brave thoughts a day, nurses an average of 12 brave thoughts a day, secretaries 6 brave thoughts a day and firemen a staggering 72 brave thoughts a day. From these results, the psychologist noted that firemen had significantly more brave thoughts each day than teachers, nurses or secretaries. She concluded that this indicated that firemen were overall much braver individuals than people in other occupations were.

Probabilistic-Structure1

At Plymouth University, the Housing Office determines which of the 10,000 students enrolled will be allowed to live on campus the following year. At Plymouth, the dormitory facilities are excellent, so there is always great demand for on-campus housing. Unfortunately, there are only enough on-campus spaces for 5000 students. The Housing Office determines who will get to live on campus by having a Housing Draw every year: every student picks a number out of a box over a 3-day period. These numbers range from 1 to 10,000. If the number is 5000 or under, the student gets to live on campus. If the number is over 5000, the student will not be able to live on campus.

On the first day of the draw, Joe talks to five people who have picked a number. Of these, four people got low numbers. Because of this, Joe suspects that the numbers in the box were not properly mixed, and that the early numbers are more favourable. He rushes over to the Housing Draw and picks a number. He gets a low number. He later talks to four people who drew their numbers on the second or third day of the draw. Three got high numbers. Joe says to himself, "I'm glad that I picked when I did, because it looks like I was right that the numbers were not properly mixed."

What do you think of Joe's reasoning? Explain.

Probabilistic - Structure 3

Bert H. has a job checking the results of an X-ray scanner of pipeline welds in a pipe factory. Overall, the X-ray scanner shows that the welding machine makes a perfect weld about 80% of the time. Of 900 welds each day, usually about 680 to 740 are perfect. Bert has noticed that on some days, all of the first 10 welds were perfect. However, Bert has also noticed that on such days, the overall number of perfect welds is usually not much better for the day as a whole than on days when the first 10 welds show some imperfections.

Why do you suppose the number of perfect welds is usually not much better on days where the first batch of welds was perfect than on other days?

Probabilistic - Structure 5

An auditor for the Inland Revenue wants to study the nature of arithmetic errors on income tax returns. She selects 4000 Social Security numbers by using random digits generated by an 'Electronic Mastermind' calculator. And for each selected social security number she checks the 1978 Tax returns thoroughly for arithmetic errors. She finds errors on a large percentage of the tax returns, often 2 to 6 errors on a single tax return. Tabulating the effect of each error separately, she finds that there are virtually the same number of errors in favour of the taxpayer as in favour of the government. Her boss objects vigorously to her assertions, saying that it is fairly obvious that people will notice and correct errors in favour of the government, but will 'overlook' errors in their own favour. Even if her figures are correct, he says, looking at a lot more returns will bear out his point.

Comment on the auditor's reasoning and her boss's contrary stand.

Objective - Structure 1

A talent scout for a professional football team attends two local games with the intention of observing carefully the talent and skill of a particular player. The player looks generally excellent. He repeatedly makes tackles worthy of the best professional players. However, in one of the games, with his team behind by 2 goals, the player is fouled while attempting to score and has the opportunity to score on a penalty free kick. The player however misses by far. The other team then goes on to score another goal and therefore wins by 3 to nil.

The scout reports that the player in question "has excellent skills, and should be recruited. He has a tendency to misplay under extreme pressure, but this will probably disappear with more experience and better coaching."

Comment on the thinking embodied in the scout's opinion that the player (a) "excellent skills" and that the player has (b) "a tendency to misplay under extreme pressure." Does the thinking behind either conclusion have any weaknesses?

Objective - Structure 3

Howard was a teacher in a secondary school in a community known for truancy and delinquency problems among its youth. Howard says of his experiences: "Usually, in a class of 35 or so kids, 2 or 3 will pull some pretty bad stunts in the first week - they'll skip a day of class, get into a scuffle with another kid, or some such thing. When that kind of thing happens, I play it down ant try to avoid calling the class' attention to it. Usually, these kids turn out to be no worse than the others. By the end of the term you'll find they haven't pulled any more stunts than the others have." Howard reasons as follows: "Some of these kids are headed toward a delinquent pattern of behaviour. When they find out nobody is very impressed, they tend to settle down."

(a) Do you agree that it is likely that the students who pull a "pretty bad stunt in the first week" are headed toward a delinquent pattern of behaviour?
(b) Do you agree that it is likely that the students who initially pull a "pretty bad stunt" turn out to be no worse than the others because they find no one is impressed with their behaviour?

Objective - Structure 5

The local education authority was urging the school board to make an expensive curriculum shift to a "back-to-basics" stress on fundamental learning skills and away from the electives and intensive immersion in specialised arts and social studies topics that had recently characterised the secondary schools in the district. He cited a study of 120 schools systems that had recently begun to emphasise the basics and 120 school systems that had a curriculum similar to the district's current one. The "back-to-basics" school systems, he said, were producing students who scored half-a-year ahead of the students in the other systems on objective tests of reading, mathematics, and science. Of the 120 "back-to-basics" school systems, 85 had shown improved skills for students in the system vs only 40 with improved skills in the 120 systems which had not changed. One of the school board members took the floor to argue against the change. In her opinion, she said, there was no compelling reason to attribute the improved student skills in the "back-to-basics" system to the specific curriculum change for two reasons: (1) School systems that make curriculum changes probably have more energetic, adventurous administrators and faculty and thus the students would learn more in those school systems no matter what the curriculum was. (2) Any change in curriculum could be expected to produce improvement in student performance because of increased faculty interest and commitment.

Comment on the reasoning of both the local education authority and the board member. On the basis of the evidence and arguments offered, do you think it is likely that the "back-to-basics" curriculum is intrinsically superior to the district's current curriculum?

Subjective- Structure 1

Gerald M. had a 3-year-old son, Timmy. He told a friend: "You know, I've never been much for sports, and I think Timmy will turn out the same. A couple of weeks ago, an older neighbour boy was tossing a ball to him, and he could catch it and throw it all right, but he just didn't seem interested in it. Then the other day, some kids his age were kicking a little soccer ball around. Timmy could do it as well as the other, but he lost interest very quickly and started playing with some toy cars while the other kids went on kicking the ball around for another 20 or 30 minutes."

Do you agree with Gerald's reasoning that Timmy is likely not to care much for sports? Why or why not?

Subjective - Structure 3

Janice is a head nurse in a home for the aged. She says the following of her experiences: "There is a big turn over of the nursing staff here, and each year we hire 15-20 new nurses. Some of these people show themselves to be unusually warm and compassionate in the first few days. One might stay on past quitting time with a patient who's having a difficult night. Another might be obviously shaken by the distress of a patient who has just lost a spouse. I find though that over the long haul, these women turn out to be not much more concerned and caring than the others. What happens to them, I think is that they can't remain open and vulnerable without paying a heavy emotional price. They usually continue to be considerate and effective but they build up a shell."

Comment on Janice's reasoning. Do you think it is likely that she correctly identifies the nurses who are unusually warm and compassionate? Do you agree it is likely that most of the ones who are unusually warm at first later build up a shell to protect themselves emotionally?

Subjective - Structure 5

Two New Yorkers were discussing restaurants. Jane said to Ellen, "You know, most people seem to be crazy about Chinese food, but I'm not. I've been to about 20 different Chinese Restaurants, across the whole price range, and everything from bland Cantonese to spicy Szechwan and I'm really not very fond of any of it." "Oh," said Ellen, "don't jump to conclusions. I'll bet you've usually gone with a crowd of people, right?" "Yes," admitted Jane, "that's true, I usually go with half a dozen people or more from work," "Well, people that may be it!" said Ellen, "people usually go to Chinese restaurants with a crowd of people they hardly know. I know you, you're often tense and a little shy, and you're not likely to be able to relax and savour the food under those circumstances. Try going to a Chinese restaurant with just one good friend. I'll bet you'll like the food."

Comment on Ellen's reasoning. Do you think there is a good chance that if Jane went to a Chinese restaurant with one friend, she'd like the food? Why or why not?

APPENDIX A4.1 Training on the logic of the material conditional (Cheng, Holyoak, Krantz, & Nisbett (1986)

Training

In this study we are interested in how people interpret and reason about a very important type of logical statement, called the *conditional.* Even though conditional statements are really very simple, people often make errors in dealing with them. These instructions are intended to help you understand conditional statements. Read through these instructions carefully; they should help you solve some reasoning problems you will receive afterward.

A conditional statement consists of two component statements which are often joined by the conjunctions "if . . . then." The conditional statement can be expressed in the standard form ((a));

If p, then q

where it is understood that the letters "*p*" and "*q*" each represent a statement. Statement (a) means "if statement *p* is true, then statement *q* is also true." For example, let *p* stand for "It is raining," and let *q* stand for "The pavement is wet." Then (a) says "If it is raining, then the pavement is wet."

Statement (a) can be expressed in a variety of ways. We will use a horizontal bar before a letter to indicate that a statement is not true. For example, "$\neg p$" means "not *p*." One way of reformulating (a) then, is ((b));

If $\neg q$, *then* $\neg p$.

This means "if statement *q* is false, then statement p is also false. Rephrasing the above example into form (b) gives, "If the pavement is not wet, then it is not raining"

People often don't realise at first that statements (a) and (b) are equivalent (identical to each other). To understand the equivalence of statements (a) and (b), consider the circumstances under which (a) is true. The truth of *"If p, then q"* depends on the truth of *p* and *q*. The table below lists the truth values of various statements. Reading across and down the table, we see that when *p* is true and q is also true, then (a) is true (first line). When *p* is true and *q* is false, then (a) is false (second line), since (a) says that *p* implies *q*. So in order for (a) to be true when *q* is false, *p* cannot be true (comparing the second and third lines). In other words, (a) implies "If $\neg q$, then $\neg p$".

p	*q*	(a): *If p, then q*
T	T	T
T	F	F
F	F	T

To check your understanding of the conditional statement, please answer the question on the next page.

Statement (a). "*If p, then q,*" can be rephrased without changing its basic meaning. Which of the following is a correct rephrasing of (a)? Put a check next to the correct rephrasing(s) before checking the answer on the next page.

( )1. *If q, then p.*
( )2. *If* $\neg p$, *then* $\neg q$.
( )2. *If* $\neg q$, *then* $\neg p$.

Only 3 is a correct rephrasing. You should note that statement (a) "*If p, then q,*" does *not* imply ((c));

If q, then p.

It is a common error to assume that (a) implies (c). Rephrasing the example on Page 1 into form (c) gives "If the pavement is wet, then it is raining." The pavement may get wet from lawn sprinklers nearby, for instance.

It is also a common error to assume that "*If p, then q*" implies ((d));

If ¬p, then ¬q.

Rephrasing the example on Page 1 into form (d) gives "If it is *not* raining, then the pavement is *not* wet," which again does not follow from "If it is raining, then the pavement is wet," for the same reason mentioned earlier (e.g. a lawn sprinkler might have made the pavement wet even though it isn't raining).

To sum up, these are the most important facts you need to know about the conditional statement. First, statement (a) is equivalent to statement (b): "*If p, then q*" implies "*If ¬q, then ¬p.*" Second, statement (a) is *not* equivalent to either statement (c) or statement (d): "If p, then q" does *not* imply either "*If q, then p*" or "*If ¬p, then ¬q.*"

Rephrasing Exercise

This exercise will check your understanding of the conditional statement. Which of the statement(s) below follow logically from the statement, "If the tablecloth is brown, then the wall is white?" Please put a check next to the correct statement(s) below before checking the answer on the next page.

( )1. If the tablecloth is *not* brown, then the wall is *not* white.
( )2, If the wall is *not* white, then the tablecloth is *not* brown.
( )3. If the wall is white, then the tablecloth is brown.

Only 2 is correct. To see this, we can reformulate the statement into the form "*If p, then q*" by substituting *p* for "the tablecloth is brown" and *q* for "the wall is white." Then we see that 2 is in the form "*If ¬q, then ¬p*," which we saw earlier is equivalent to "If p, then q." But 1 is in the form "*If ¬p, then ¬q*," which does *not* follow from "*If p, then q*." And 3 is in the form "If q, then p," which likewise does *not* follow from "*If p, then q.*"

Below are two more rephrasing problems. When you are done, check the answer on the next page.

Statement: If the cube is plastic, then the sphere is metallic.
( )1. If the sphere is metallic, then the cube is plastic.
( )2. If the cube is *not* plastic, then the sphere is *not* metallic
( )3. If the sphere is *not* metallic, then the cube is *not* plastic.

Statement: If the beach is white, then the music is slow.
( )1. If the beach is *not* white, then the music is *not* slow.
( )2. If the music is slow, then the beach is white.
( )3. If the music is *not* slow, then the beach is *not* white.

Answer to rephrasing problems: Only 3 is correct in the above two problems.

Following are two examples that illustrate how the conditional statement is used to solve problems. Please think carefully and solve the problems before checking the answers beneath them.

Example 1

You have a set of four stamps with either a red or blue star, or a red or blue triangle on it. You are told that if there is a blue star on one side, then there is a red triangle on the other. Each of the stamps are laid side by side face upwards. Which card/cards would you need to turn over to check that the rule is true?

(a)	(b)	(c)	(d)
Red Star	**Blue Star**	**Red Triangle**	**Blue Triangle**
turn ( )	turn ( )	turn ( )	turn ( )

Answer to Example One

To reformulate the rule into the form *"If p, then q,"* we substitute *p* for "If there is a blue star on one side," and *q* for "there is a red triangle on the other." Since alternative (b) corresponds to *p*, we clearly have to check whether *q* follows. Since (d) corresponds to $\neg q$ and *"If $\neg q$, then $\neg p$"* is equivalent to *"If p, then q,"* we have to check whether $\neg p$ follows.

But we need not turn over stamp (a) which corresponds to $\neg p$, and as said earlier *"If p, then q"* does not imply *"If $\neg p$, then $\neg q$."* Nor do we need to turn over stamp (c), which translated into *q*. As you learned earlier, *"If p, then q"* is not equivalent to *"If q, then p"* accordingly, the other side of stamp (c) is irrelevant to the truth of the rule.

Example 2

You have four postcards. If there is a picture of a beach on one side then there is a ball on the other. Each of the cards are face-up on a table. Which postcard(s) would you turn over to check that the rule is true?

(a)	(b)	(c)	(d)
Ball	**Frisbee**	**Beach**	**Hills**
turn ( )	turn ( )	turn ( )	turn ( )

Answer to Example Two

To reformulate the above conditional statement into the form *"If p, then q,"* we substitute p for " If there is a picture of a beach," and q for "there is a ball." Since card (c) translates into p, we clearly have to check whether q follows. And since card (b) translates into $\neg q$, and *"If $\neg q$, then $\neg p$"* is equivalent to *"If p, then q,"* we have to check whether $\neg p$ follows.

But we need not turn over card (a), since it corresponds to q, and *"If q, then p"* does not follow from *"If p, then q."* Similarly, we need not turn over card (d), which corresponds to $\neg p$, a condition whose implications are irrelevant to the truth of *"If p, then q."*

APPENDIX A4.2 Obligation Schema Training (Cheng, Holyoak, Krantz, & Nisbett (1986)

Training

As you know, an obligation arises whenever it is the case that certain circumstances or situations create an obligation to perform some action. Obligations can often be stated in an *"If. . . then"* form. For example, the following regulation specifies an obligation: "If a student is a psychology major, then the student must take an introductory psychology course." More generally, if we call the initial situation *I* and the action *C*, an obligation has the form, "If *I* arises, then *C* must be done." In our first example, *I* is " being a psychology major," and *C* is "taking an introductory psychology course."

In order to assess whether an obligation is being satisfied, we need to consider the four possible situations that might arise. These are

1. *I* occurs.
2. *I* doesn't occur.
3. *C* is done.
4. *C* is not done.

Corresponding to each of these possible situations is a rule related to the fulfillment of the obligation. These rules are the following:

1. If *I* occurs, then it is obligatory to do *C*. Clearly, if *I* arises then failure to take the required action would constitute a violation of the obligation. To use our example, if a student is a psychology major, then that student must take an introductory psychology course.
2. If *I* does not occur, then the obligation does not arise. Consequently, *C* need not be done, although the person may do *C* anyway. For example, if a student is not a psychology major the student is not obliged to take an introductory psychology course. It may be permissible, however, for an English major to take an introductory psychology course. But in any case, the basic obligation is simply irrelevant if the student is not a psychology major.
3. If *C* is done, then the obligation is certainly not violated, regardless of whether or not *I* has occurred. If *I* did occur, then the obligation is satisfied. If *I* didn't occur, then the obligation didn't even arise (Rule 2). For example, if we know a student has taken an introductory psychology course, we can be sure the obligation has not been violated: Either the student was a psychology major, and hence fulfilled the obligation, or the student was not a psychology major, in which case the obligation didn't arise.
4. If *C* has not been done, then *I* must not have occurred. This is because if *I* had occurred, then the failure to do *C* would constitute a violation of the obligation. Thus, if a student has not taken an introductory psychology course, the student must not be a psychology major, or else the obligation will have been violated.

If you understand the above four rules, you should find it very easy to assess whether or not an obligation is being met. Note that there are only two situations in which it is possible for an obligation to be violated: When *I* occurs (and *C* is not done) (Rule 1), and when *C* is not done (and *I* occurs) (Rule 4). In the other two situations the obligation can't be violated. These are the cases in which *I* doesn't occur (in which case the obligation doesn't arise) (Rule 2), and in which *C* is done (in which case the obligation will have been met if it arose) (Rule 3).
Following are two examples that illustrate how these rules may be used to solve problems. Please think carefully and solve the problems before checking the answers on the following pages.

Example 1

Imagine you are a parent asking your child to help you tidy up around the house. You have learnt that just asking nicely doesn't always get the desired response therefore you have set some rules. One such rule is:

"If you want to watch television, then you must tidy your room first."

This rule may or may not be in effect. You have to make sure that this rule is being followed. The four cards below represent information about whether your child tidies up and whether they watch television. Each card has two pieces of information on it. On one side of the card there is information about the room being tidied or not, and on the other there is information about television watching. For two of the cards you can see whether or not your child tidies their room, but not whether they watched television, and on the other two you can see whether or not they get to watch television, but not whether they tidied their room. Your task is to determine whether or not your child is following this rule. Indicate which card(s) you would need to turn over to decide whether or not the rule is being broken.

(a)	(b)	(c)	(d)
Tidies up	**Doesn't tidy up**	**Does watch television**	**Does not watch television**
turn ( )	turn ( )	turn ( )	turn ( )

Answer to Example 1

The correct answer is to circle the 'Watches television' and the 'Doesn't tidy up' cards (c and b). The rule states that if the child is to 'watch television', then they must 'tidy their room first'. That is, if *I* occurs, then it is obligatory to do *C*, therefore the 'Watches television' card is turned to check that 'Tidies up' is on the other side. If 'watching television' arises with 'Doesn't tidy up' on the other side then this would constitute a violation of the obligation. The 'Doesn't tidy up' card is also turned to check that 'Doesn't watch television' is on the other side. In other words, if *C* has not been done, then *I* must not have occurred. If 'Watches television' is on the other side then this again would be a violation of the obligation.

It is not necessary to look on the other sides of the 'Tidies up' card (a) or the 'Doesn't watch television' cards (d). If the child 'Tidies up' the obligation has not been violated whether they watch television or not. Also if the child 'Doesn't watch television', then the obligation doesn't arise anyway, they don't have to tidy their room.

Example 2

As a teacher at a secondary school it is your job to make sure all assignments for your subject are in before the end of term. One of the rules you have set is;

"If you want to go to the school dance, then you must complete all your assignments."

This rule may or may not be in effect. You have to make sure that this rule is being followed. There are four cards below, each depicting a pupil at the school. Each card shows two pieces of information. On one side whether or not they have completed their assignments and on the other side whether they went to the school dance or not. For two of the pupils you can see if they completed their assignments, but not whether they went to the school dance, and for the other two you can see whether they attended the dance, but not whether they completed their assignments.

You have to determine whether or not the rule above is being broken. Indicate which card(s) you would need to turn over to decide whether or not the rule is being broken.

(a)	(b)	(c)	(d)
No School dance	**Essays complete**	**School dance**	**Essays not Complete**
turn ( )	turn ( )	turn ( )	turn ( )

Answer to Example 2

The corrects answer is to indicate the 'School dance' and the 'Essays not complete' cards (c and d). The rule states that if the pupils go to the 'School dance' then they must 'complete their essays', therefore the 'school dance' card is turned to check that this has been done. If the required action has not been taken then this would constitute a violation of the obligation (If *I* occurs, then it is obligatory to do *C*.). The 'Essays not complete' card is also turned to check that this pupil has not gone to the school dance. In other words, if *C* has not been done, then *I* must not have occurred, if it has (e.g. pupil attended school dance) then again this would be a violation of the obligation.

There is no point turning the 'No school dance' card (a) as it is irrelevant whether they completed their assignments or not (If *I* does not occur, then the obligation does not arise). It is also unnecessary to turn the 'Essays complete' card (b) as the obligation can not be violated, whether or not the pupil attends the school dance.

APPENDIX A4.3 Conditional reasoning problems

Instructions

For this booklet of logical problems your task is to decide which of the three conclusions that follow the premises are valid, i.e. those which logically have to follow given that the premises are true.

Each problem consists of two statements. An example of one of the problems is shown below:-

If there is a tree, then there is a balloon.
There is a tree.
Which of the following conclusions is valid?
(a) There is a balloon
(b) There is not a balloon
(c) It is uncertain whether or not there is a balloon.

If you think the logical conclusion is that there is a balloon, then circle (a).

You should assume that the information is true. Your task is to decide which of the conclusions follow logically from the statements. Please respond with just one option per problem.

Please do not turn back and forth from one problem to another once you have started. You must not make notes or draw diagrams of any kind to help you in this task.

If there is a 7, then there is a D.
There is a 7.
Which of the following conclusions is valid?
(a) There is a D
(b) There is not a D
(c) It is uncertain whether or not there is a D.

If there is a 4, then there is a P.
There is a P.
Which of the following conclusions is valid?
(a) There is a 4
(b) There is not a 4
(c) It is uncertain whether or not there is a 4.

If there is a 9, then there is T.
There is not a 9.
Which of the following conclusions is valid?
(a) There is a T
(b) There is not a T
(c) It if uncertain whether or not there is a T.

If there is an 8, then there is a K.
There is not a K.
Which of the following conclusions is valid?
(a) There is an 8
(b) There is not an 8
(c) It is uncertain whether or not there is an 8.

If there is a square, then there is a circle.
There is a square.
Which of the following conclusions is valid?
(a) There is a circle
(b) There is not a circle
(c) It is uncertain whether or not there is a circle

If there is a rectangle, then there is a triangle.
There is a triangle.
Which of the following conclusions is valid?
(a) There is a rectangle
(b) There is not a rectangle
(c) It is uncertain whether or not there is a rectangle.

If there is a triangle, then there is a square.
There is not a triangle.
Which of the following conclusions is valid?
(a) There is a square
(b) There is not a square
(c) It is uncertain whether or not there is a square

If there is a circle, then there is a rectangle.
There is not a rectangle.
Which of the following conclusions is valid?
(a) There is a circle
(b) There is not a circle
(c) It is uncertain whether or not there is a circle.

If there is a hammer, then there is a wrench.
There is a hammer.
Which of the following conclusions is valid?
(a) There is a wrench
(b) There is not a wrench
(c) It is uncertain whether or not there is a wrench.

If there is a chisel, then there is a saw.
There is a saw.
Which of the following conclusions is valid?
(a) There is a chisel
(b) There is not a chisel
(c) It is uncertain whether or not there is a chisel.

If there is a screwdriver, then there is a mallet.
There is not a screwdriver.
Which of the following conclusions is valid?
(a) There is a mallet
(b) There is not a mallet
(c) It is uncertain whether there is a mallet or not.

If there is a drill, then there is a shovel.
There is not a shovel.
Which of the following conclusions is valid?
(a) There is a drill
(b) There is not a drill
(c) It is uncertain whether there is a drill or not.

If there is an apple, then there is a banana.
There is an apple.
Which of the following conclusions is valid?
(a) There is a banana
(b) There is not a banana
(c) It is uncertain whether there is a banana or not.

If there is a grapefruit, then there is a pineapple.
There is a pineapple.
Which of the following conclusions is valid?
(a) There is a grapefruit
(b) There is not a grapefruit
(c) It is uncertain whether there is a grapefruit or not.

If there is an orange, then there is a peach.
There is not an orange.
Which of the following conclusions is valid?
(a) There is a peach
(b) There is not a peach
(c) It is uncertain whether or not there is a peach.

If there is a plum, then there is a cherry.
There is not a cherry.
Which of the following conclusions is valid?
(a) There is a plum
(b) There is not a plum
(c) It is uncertain whether or not there is a plum.

APPENDIX A5.1 Permission Training (Klaczynski & Laipple, 1993)

Imagine you work in the Faculty of Science. Part of your job involves collating all the information required about students who are graduating each year. You have to make sure that students who want to graduate in the December have completed their course by September, and certain rules have to be followed to ensure a student is eligible. One of these rules is:

"If a student graduates, then that student must have completed all their modules."

You have to make sure that this rule is being followed. The four cards below represent students at your college. Each card has two pieces of information on it. On one side of each card has whether a student has graduated and on the other side whether they completed their modules. For two of the cards you can see whether they graduated, but you cannot see if they completed their modules. For the other two cards, you can see whether they completed all their modules, but you cannot see whether they graduated. Indicate which card(s) you would need to turn over to decide whether or not the rule is being violated.

(a)	(b)	(c)	(d)
Not Graduate	**Graduate**	**Modules not Complete**	**Modules Complete**
turn ( )	turn ( )	turn ( )	turn ( )

Explanation

The correct answer is to turn over both the 'Graduated' and the 'Modules not Completed' cards. The rule states that if a student graduates, then that student must have completed all their modules.

The 'Graduated' card must be turned over because whether or not the rule is being followed is determined by whether or not that student has completed all their modules. If it states that the student has not completed all their modules then the rule is being violated. Likewise, the 'Modules not Completed' card must be turned over because, according to the rule, this student must not have graduated. If a student has not completed all their modules but has graduated then the rule is being violated. In order to see if a student has graduated without completing all their modules, one must turn this card over.

The 'Not Graduated' card does not have to be turned over because whether or not the student completed their modules could be on the other side of the card and the rule would still be followed; that is, whether or not a student completed all their modules, if they didn't graduate it doesn't affect the rule, so the information on the other side of this card is irrelevant. The 'Modules Completed' card does not have to be turned over either. Because the rule is not reversible, a student who has completed all their modules does not have to graduate; either Graduated or Not Graduated could be on the other side of this card and the rule would still be followed.

APPENDIX A5.1 cont.

Imagine you work as a Receptionist in a Doctor's surgery. Part of your job involves making sure all the patients are up to date with their vaccinations. You also have to make sure that patients who go abroad for their holidays are receiving the inoculations they require for the destinations they are travelling to. The Doctor has given you a set of guidelines and rules that should be adhered to. One of these rules is:

"If a patient is travelling to Africa, then that patient must be vaccinated against Yellow Fever."

You have to make sure that this rule is being followed. The four cards below represent patients from the surgery. Each card has two pieces of information on it. On one side of each card is whether a patient has been to Africa and on the other side whether they have been vaccinated. For two of the cards you can see whether they travelled to Africa, but you cannot see if they were vaccinated. For the other two cards, you can see whether they were vaccinated, but you cannot see whether they travelled to Africa. Indicate which card(s) you would need to turn over to decide whether or not the rule is being violated.

(a)	(b)	(c)	(d)
Spain	**Africa**	**Vaccinated**	**Not Vaccinated**
turn ( )	turn ( )	turn ( )	turn ( )

<u>**Explanation**</u>

The correct answer is to turn over both the 'Africa' and the 'Not vaccinated' cards. The rule states that if a patient travels to Africa, then that patient must be vaccinated against Yellow Fever.

The 'Africa' card must be turned over because whether or not the rule is being followed is determined by whether or not that patient has been vaccinated. If it states that the patient has not been vaccinated then the rule is being violated. Likewise, the 'Not vaccinated' card must be turned over because, according to the rule, this patient must not have travelled to Africa. If a patient has not received their vaccination but has travelled to Africa then the rule is being violated. In order to see if a patient has travelled to Africa without receiving their vaccination, one must turn this card over.

The 'Spain' card does not have to be turned over because whether or not the patient was vaccinated could be on the other side of the card and the rule would still be followed; that is, whether or not a patient received their vaccination, if they didn't travel to Africa it doesn't affect the rule, so the information on the other side of this card is irrelevant. The 'Vaccinated' card does not have to be turned over either. Because the rule is not reversible, a patient who has received their vaccination does not have to travel to Africa; either Africa or Spain could be on the other side of this card and the rule would still be followed.

Imagine you are a parent asking your child to help you tidy up around the house. You have learnt that just asking nicely doesn't always get the desired response, therefore you have set some rules. One of these rules is:

"If you want to watch television, then you must tidy your room first."

You have to make sure that this rule is being followed. The four cards below represent information about whether your child tidies up or not and whether or not they watch television. Each card has two pieces of information on it. On one side of each card you can see whether or not your child tidies their room, and on the other whether they watched television. For two of the cards you can see whether they watched television, but you cannot see whether they tidied their room. For the other two cards, you can see whether they tidied their room but not whether they watched television. Indicate which card(s) you would need to turn over to decide whether or not the rule is being violated.

(a)	(b)	(c)	(d)
Tidies up	**Doesn't tidy up**	**Watches television**	**Doesn't watch television**
turn ( )	turn ( )	turn ( )	turn ( )

Explanation

The correct answer is to turn over both the 'Watches television' and the 'Doesn't tidy up' cards. The rule states that if the child wants to watch television, then the child must tidy their room first.

The 'Watches television' card must be turned over because whether or not the rule is being followed is determined by whether or not the child tidies their room. If it states that the child doesn't tidy their room then the rule is being violated. Likewise, the 'Doesn't tidy up' card must be turned over because, according to the rule, the child must not have watched television. If the child has not tidied their room but has watched television then the rule is being violated. In order to see if the child has watched television without tidying their bedroom, one must turn this card over.

The 'Doesn't watch television' card does not have to be turned over because whether or not the child tidied their room could be on the other side of the card and the rule would still be followed; that is, whether or not the child tidied their room, if they didn't watch television it doesn't affect the rule, so the information on the other side of this card is irrelevant. The 'Tidies up' card does not have to be turned over either. Because the rule is not reversible, a child who has tidied their room does not have to watch television; either watches television or doesn't watch television could be on the other side of this card and the rule would still be followed.

Imagine you are the captain of the England Athletics team. It is your job to decide who represents England in different competitions. You want the decisions to be fair and everyone to have a chance on being in the team, but you don't want complaints of unfairness after the decisions have been made, therefore you have set some rules. One of these rules is:

"If an athlete is to represent England, then they must win all the heats first."

You have to make sure that this rule is being followed. The four cards below represent athletes competing for places on the team. Each card has two pieces of information on it. On one side of each card is whether an athlete has represented England and on the other side whether they won all their heats. For two of the cards you can see whether they represented England but you cannot see if they won all their heats. For the other two cards, you can see whether they won all their heats but you cannot see whether they represented England. Indicate which card(s) you would need to turn over to decide whether or not the rule is being violated.

(a)	(b)	(c)	(d)
Represents England	**Won all Heats**	**Didn't Win all Heats**	**Didn't represent England**
turn ( )	turn ( )	turn ( )	turn ( )

Explanation

The correct answer is to turn over both the 'Represents England' and the 'Didn't win all heats' cards. The rule states that if an athlete is to represent England, then they must win all the heats first.

The 'Represents England' card must be turned over because whether or not the rule is being followed is determined by whether or not the athlete wins the heats first. If it states that the athlete doesn't win the heats then the rule is being violated. Likewise, the 'Didn't win all heats' card must be turned over because, according to the rule, the athlete must not have represented England. If the athlete has not won all the heats but has represented England then the rule is being violated. In order to see if the athlete has represented England without winning all the heats, one must turn this card over.

The 'Didn't represent England' card does not have to be turned over because whether or not the athlete won all the heats could be on the other side of the card and the rule would still be followed; that is, whether or not the athlete won all the heats, if they didn't represent England it doesn't affect the rule, so the information on the other side of this card is irrelevant. The 'Won all heats' card does not have to be turned over either. Because the rule is not reversible, an athlete who has won all the heats does not have to represent England; either represents England or didn't represent England could be on the other side of this card and the rule would still be followed.

APPENDIX A5.2 Abstract Training (Klaczynski, Gelfand, & Reese, 1989)

The boxes below represent four cards lying on a table. Each card has two pieces of information on it. On one side of each card, there is a fraction, and on the other side of each card, there is a chemical element. For two of these cards, you can see the fraction that is on one side of that card, but you cannot see the chemical element that is on the other side of that card. For the other two cards, you can see the chemical element that is on one side of the card, but you cannot see the fraction that is on the other side. Your friend tells you that the following rule applies to the cards:

"If 1/2 is on one side of a card, then Helium must be on the other side of the card."

Your task is to indicate the card or cards you would need to turn over to see whether this rule is true or false.

(a)	(b)	(c)	(d)
3/4	**1/2**	**Helium**	**Oxygen**
turn ( )	turn ( )	turn ( )	turn ( )

<u>Explanation</u>

The correct answer is to turn over both the '1/2' and the 'Oxygen' cards. The rule states that if there is a 1/2 on one side of a card, then Helium must be on the other side of that card.

The '1/2' card must be turned over because the truth of the rule is determined by the chemical element on the other side of this card. If the element on the other side of this card is any element other than helium, then the rule cannot be true. Likewise, the 'Oxygen' card must be turned over because, according to the rule, Oxygen cannot be on the other side of a card that has 1/2 on it. If a card with Oxygen on one side has 1/2 on the other side of it, the rule cannot be true. In order to see if the 'Oxygen' card has 1/2 on the other side of it, one must turn this card over.

The '3/4' card does not have to be turned over because any chemical could be on the other side of this card and the rule could still be true; that is, the chemical that appears on the other side of this card does not affect the truth of the rule, so the information on the other side of this card is irrelevant. The 'Helium' card does not have to be turned over either. Because the rule is not reversible, a card with Helium on one side does not have to have 1/2 on the other side of it; either 1/2 or 3/4 could be on the other side of this card and the rule could still be true.

The boxes below represent four cards lying on a table. Each card has two pieces of information on it. On one side of each card, there is an occupation, and on the other side of each card, there is a hair colour. For two of these cards, you can see the occupation that is on one side of that card, but you cannot see the hair colour that is on the other side of that card. For the other two cards, you can see the hair colour that is on one side of the card, but you cannot see the occupation that is on the other side. Your friend tells you that the following rule applies to the cards:

"If Doctor is on one side of a card, then Blonde must be on the other side of the card."

Your task is to indicate the card or cards you would need to turn over to see whether this rule is true or false.

(a)	(b)	(c)	(d)
Brunette	**Blonde**	**Postman**	**Doctor**
turn ( )	turn ( )	turn ( )	turn ( )

Explanation

The correct answer is to turn over both the 'Doctor' and the 'Brunette' cards. The rule states that is there is a Doctor on one side of a card, then Blonde must be on the other side of that card.

The 'Doctor' card must be turned over because the truth of the rule is determined by the hair colour on the other side of this card. If the hair colour on the other side of this card is any colour other than blonde, then the rule cannot be true. Likewise, the 'Brunette' card must be turned over because, according to the rule, brunette cannot be on the other side of a card that has Doctor on it. If a card with brunette on one side has Doctor on the other side of it, the rule cannot be true. In order to see if the 'Brunette' card has Doctor on the other side of it, one must turn this card over.

The 'Postman' card does not have to be turned over because any hair colour could be on the other side of this card and the rule could still be true; that is, the hair colour that appears on the other side of this card does not affect the truth of the rule, so the information on the other side of this card is irrelevant. The 'Blonde' card does not have to be turned over either. Because the rule is not reversible, a card with Blonde on one side does not have to have Doctor on the other side of it; either Doctor or Postman could be on the other side of this card and the rule could still be true.

The boxes below represent four cards lying on a table. Each card has two pieces of information on it. On one side of each card, there is a colour, and on the other side of each card there is a religion. For two of these cards, you can see the colour that is on one side of that card, but you cannot see the religion that is on the other side of the card. For the other two cards, you can see the religion that is on one side of the card, but you cannot see the colour that is on the other side. Your friend tells you that the following rule applies to the cards:

"If Pink is on one side of a card, then Protestant must be on the other side."

Your task is to indicate the card or cards you would need to turn over to see whether this rule is true or false.

(a)	(b)	(c)	(d)
Catholic	**Purple**	**Pink**	**Protestant**
turn ( )	turn ( )	turn ( )	turn ( )

Explanation

The correct answer is to turn over both the 'pink' and the 'Catholic' cards. The rule states that if there is a pink on one side of a card, then Protestant must be on the other side of that card.

The 'pink' card must be turned over because the truth of the rule is determined by the religion on the other side of this card. If the religion on the other side of this card is any religion other than Protestant, then the rule cannot be true. Likewise, the 'Catholic' card must be turned over because, according to the rule, Catholic cannot be on the other side of a card that has pink on it. If a card with Catholic on one side has pink on the other side of it, the rule cannot be true. In order to see if the 'Catholic' card has pink on the other side of it, one must turn this card over.

The 'Purple' card does not have to be turned over because any religion could be on the other side of this card and the rule could still be true; that is, the religion that appears on the other side of this card does not affect the truth of the rule, so the information on the other side of this card is irrelevant. The 'Protestant' card does not have to be turned over either. Because the rule is not reversible, a card with Protestant on one side does not have to have pink on the other side of it; either pink or purple could be on the other side of this card and the rule could still be true.

The boxes below represent four cards lying on a table. Each card has two pieces of information on it. On one side of each card, there is a season, and on the other side of each card, there is a type of aeroplane. For two of these cards, you can see the season that is on one side of that card, but you cannot see the type of aeroplane on the other side of that card. For the other two cards, you can see the type of aeroplane that is on one side of the card, but you cannot see the season that is on the other side. Your friend tells you that the following rule applies to the cards:

"If Winter is on one side of a card, then Jet must be on the other side of the card."

Your task is to indicate the card or cards you would need to turn over to see whether this rule is true or false.

(a)	(b)	(c)	(d)
Biplane	**Winter**	**Jet**	**Spring**
turn ( )	turn ( )	turn ( )	turn ( )

Explanation

The correct answer is to turn over both the 'Winter' and the 'Biplane' cards. The rule states that if there is a Winter on one side of a card, then Jet must be on the other side of that card.

The 'Winter' card must be turned over because the truth of the rule is determined by the type of aeroplane on the other side of this card. If the aeroplane on the other side of this card is any aeroplane other than Jet, then the rule cannot be true. Likewise, the 'Biplane' card must be turned over because, according to the rule, Biplane cannot be on the other side of a card that has Winter on it. If a card with Biplane on one side has Winter on the other side of it, the rule cannot be true. In order to see if the 'Biplane' card has Winter on the other side of it, one must turn this card over.

The 'Spring' card does not have to be turned over because any type of aeroplane could be on the other side of this card and the rule could still be true; that is, the type of aeroplane that appears on the other side of this card does not affect the truth of the rule, so the information on the other side of this card is irrelevant. The 'Jet' card does not have to be turned over either. Because the rule is not reversible, a card with Jet on one side does not have to have Winter on the other side of it; either Winter or Spring could be on the other side of this card and the rule could still be true.

APPENDIX B1 – Experiment 1

Table 1.1 Repeated measures ANOVA to test for differences dependent on problem type in statistical reasoning

	SS	Degr. of	MS	F	p
Intercept	682.6667	1	682.6667	300.4551	0.000000
Error	111.3333	49	2.2721		
Problem type	7.4533	2	3.7267	3.0296	0.052883
Error	120.5467	98	1.2301		

Table 1.2 Post-hoc comparison

	R1	neutral	untypical	typical
1	Neutral			
2	inconsistent	0.016724		
3	consistent	0.128548	0.369456	

Table 1.3 Repeated measures ANOVA to test for differences dependent on problems type in experiment evaluation reasoning

	SS	Degr. of	MS	F	p
Intercept	1003.627	1	1003.627	219.1780	0.000000
Error	224.373	49	4.579		
Problem type	6.813	2	3.407	2.2378	0.112115
Error	149.187	98	1.522		

Table 1.4 2 x 2 (Logic x Belief) ANOVA performed on syllogistic reasoning

	SS	Deg. of	MS	F	p
Intercept	1479.680	1	1479.680	1009.528	0.000000
Error	71.820	49	1.466		
LOGIC	60.500	1	60.500	39.007	0.000000
Error	76.000	49	1.551		
BELIEF	16.820	1	16.820	17.286	0.000129
Error	47.680	49	0.973		
LOGIC*BELIEF	25.920	1	25.920	30.545	0.000001
Error	41.580	49	0.849		

Table 1.5 T-test to test difference between deontic and arbitrary selection task responding

	Mean	Std.Dv.	N	Diff.	Std.Dv.	t	df	p
arbitrary	1.420000	1.715357						
deontic	2.900000	1.432138	50	-1.48000	1.681108	-6.22517	49	0.000000

Table 2.1 ANOVA performed on the statistical reasoning problems dependent on condition

	SS	Degr. of	MS	F	p
Intercept	1095.200	1	1095.200	421.4792	0.000000
Condition	102.756	1	102.756	39.5447	0.000000
Error	150.711	58	2.598		
Problem Type	13.233	2	6.617	6.9425	0.001420
Prob type x Condition	7.544	2	3.772	3.9580	0.021736
Error	110.556	116	0.953		

Table 2.2 Follow up LSD test performed on problem type

	R1	neutral	consistent	inconsistent
1	Neutral			
2	Consistent	0.925661		
3	Inconsistent	0.001894	0.001403	

Table 2.3 Planned comparisons for the interaction in statistical reasoning under the control condition – Neutral and Consistent

	Sum of	Degr. of	Mean	F	p
M1	1.06667	1	1.066667	1.286655	0.261332
Error	48.08333	58	0.829023		

Table 2.4 Planned comparisons for the interaction in statistical reasoning under the control condition – Neutral and Inconsistent

	Sum of	Degr. of	Mean	F	p
M1	17.06667	1	17.06667	17.71838	0.000090
Error	55.86667	58	0.96322		

Table 2.5 Planned comparisons for the interaction in statistical reasoning under the control condition – Consistent and Inconsistent

	Sum of	Degr. of	Mean	F	p
M1	9.60000	1	9.600000	8.997576	0.003979
Error	61.88333	58	1.066954		

Table 2.6 Planned comparisons for the interaction in statistical reasoning under the training condition – Neutral and Consistent

	Sum of	Degr. of	Mean	F	p
M1	1.35000	1	1.350000	1.628423	0.207008
Error	48.08333	58	0.829023		

Table 2.7 Planned comparisons for the interaction in statistical reasoning under the training condition – Neutral and Inconsistent

	Sum of	Degr. of	Mean	F	p
M1	0.06667	1	0.066667	0.069212	0.793419
Error	55.86667	58	0.963218		

Table 2.8 Planned comparisons for the interaction in statistical reasoning under the training condition – Consistent and Inconsistent

	Sum of	Degr. of	Mean	F	p
M1	2.01667	1	2.016667	1.890116	0.174477
Error	61.88333	58	1.066954		

Table 2.9 ANOVA performed on the evidence evaluation scale ratings

	SS	Degr. of	MS	F	p
Intercept	26523.47	1	26523.47	1304.224	0.000000
Condition	20.67	1	20.67	1.017	0.317538
Error	1179.52	58	20.34		
Problem type	842.41	2	421.21	54.399	0.000000
Problem type x Condition	50.74	2	25.37	3.277	0.041270
Error	898.18	116	7.74		

Table 2.10 Post-hoc analysis on evidence evaluation ratings dependent on problem type

	PROBTYPE	neutral	consistent	inconsistent
1	neutral		0.600654	0.000000
2	consistent	0.600654		0.000000
3	inconsistent	0.000000	0.000000	

Table 2.11 ANOVA performed on persuasiveness ratings

	SS	Degr. of	MS	F	p
Intercept	23324.45	1	23324.45	1733.271	0.000000
Condition	6.05	1	6.05	0.450	0.505194
Error	780.50	58	13.46		
Problem type	863.63	2	431.82	53.692	0.000000
Problem type x Condition	35.43	2	17.72	2.203	0.115087
Error	932.93	116	8.04		

Table 2.12 Post-hoc analysis performed on problem type

	Problem type	neutral	consistent	inconsistent
1	Neutral			
2	Consistent	0.320419		
3	Inconsistent	0.000000	0.000000	

Table 2.13 ANOVA performed on Experiment Evaluation reasoning

	SS	Degr. of	MS	F	p
Intercept	1065.800	1	1065.800	301.4454	0.000000
Condition	1.800	1	1.800	0.5091	0.478389
Error	205.067	58	3.536		
Problem type	196.300	2	98.150	68.1217	0.000000
Problem type x condition	7.900	2	3.950	2.7415	0.068652
Error	167.133	116	1.441		

Table 2.14 Post-hoc analysis on problem type for experiment evaluation reasoning

	PROB	neutral	consistent	inconsistent
1	neutral			
2	consistent	0.000000		
3	inconsistent	0.000002	0.000000	

2.15 ANOVA performed on the evidence evaluation ratings

	SS	Degr. of	MS	F	p
Intercept	27232.20	1	27232.20	660.8835	0.000000
Condition	115.20	1	115.20	2.7957	0.099903
Error	2389.93	58	41.21		
Problem type	790.83	2	395.42	44.3172	0.000000
Problem type x condition	30.83	2	15.42	1.7279	0.182205
Error	1035.00	116	8.92		

2.16 Post-hoc analysis on evidence evaluation ratings

	PROB	neutral	consistent	inconsistent
1	Neutral			
2	consistent	0.000000		
3	inconsistent	0.000629	0.000000	

2.17 ANOVA performed on the experiment validity ratings

	SS	Degr. of	MS	F	p
Intercept	25872.02	1	25872.02	874.3151	0.000000
Condition	341.69	1	341.69	11.5470	0.001232
Error	1716.29	58	29.59		
Problem type	568.68	2	284.34	47.2285	0.000000
Problem type x condition	46.94	2	23.47	3.8987	0.022977
Error	698.38	116	6.02		

2.18 Post-hoc analysis on validity ratings

	R1	neutral	consistent	inconsistent
1	neutral		0.000001	0.000016
2	consistent	0.000001		0.000000
3	inconsistent	0.000016	0.000000	

2.19 ANOVA performed on the Fong et al. problems

	SS	Degr. of	MS	F	p
Intercept	605.0000	1	605.0000	213.6450	0.000000
Condition	74.7556	1	74.7556	26.3986	0.000003
Error	164.2444	58	2.8318		
Problem type	104.6333	2	52.3167	39.9200	0.000000
Problem type x condition	3.3444	2	1.6722	1.2760	0.283045
Error	152.0222	116	1.3105		

2.20 Post-hoc analysis on problem type

	Problem type	objective	probabilistic	subjective
1	Objective			
2	Probabilistic	0.000000		
3	Subjective	0.012019	0.000000	

Table 3.1 ANOVA performed on the LLN problems

	SS	Degr. of	MS	F	p
Intercept	2430.000	1	2430.000	546.5615	0.000000
Condition	319.200	2	159.600	35.8976	0.000000
Error	386.800	87	4.446		
Problem type	37.956	2	18.978	13.0899	0.000005
Problem type x condition	17.778	4	4.444	3.0655	0.017980
Error	252.267	174	1.450		

Table 3.2 Post-hoc analysis on Condition

	Condition	control	instruction	training
1	control			
2	instruction	0.526260		
3	training	0.000000	0.000000	

Table 3.3 Post-hoc analysis on Problem type

	Problem type	consistent	inconsistent	neutral
1	consistent			
2	inconsistent	0.002293		
3	neutral	0.049180	0.000001	

Table 3.4 Planned comparisons performed on the interaction between condition and problem type under the control group – consistent and inconsistent

	Sum of	Degr. of	Mean	F	p
M1	19.2667	1	19.26667	11.97571	0.000837
Error	139.9667	87	1.60881		

Table 3.5 Planned comparisons performed on the interaction between condition and problem type under the control group – consistent and neutral

	Sum of	Degr. of	Mean	F	p
M1	1.6667	1	1.666667	1.195712	0.277199
Error	121.2667	87	1.393870		

Table 3.6 Planned comparisons performed on the interaction between condition and problem type under the control group – inconsistent and neutral

	Sum of	Degr. of	Mean	F	p
M1	32.2667	1	32.26667	23.95903	0.000004
Error	117.1667	87	1.34674		

Table 3.7 Planned comparisons performed on the interaction between condition and problem type under the instruction group – inconsistent and consistent

	Sum of	Degr. of	Mean	F	p
M1	7.3500	1	7.350000	4.568588	0.035369
Error	139.9667	87	1.608812		

Table 3.8 Planned comparisons performed on the interaction between condition and problem type under the instruction group – neutral and consistent

	Sum of	Degr. of	Mean	F	p
M1	2.4000	1	2.400000	1.721825	0.192910
Error	121.2667	87	1.393870		

Table 3.9 Planned comparisons performed on the interaction between condition and problem type under the instruction group – neutral and inconsistent

	Sum of	Degr. of	Mean	F	p
M1	18.1500	1	18.15000	13.47696	0.000416
Error	117.1667	87	1.34674		

Table 3.10 Planned comparisons performed on the interaction between condition and problem type under the training group – consistent and inconsistent

	Sum of	Degr. of	Mean	F	p
M1	0.4167	1	0.416667	0.258990	0.612103
Error	139.9667	87	1.608812		

Table 3.11 Planned comparisons performed on the interaction between condition and problem type under the training group – neutral and consistent

	Sum of	Degr. of	Mean	F	p
M1	1.6667	1	1.666667	1.195712	0.277199
Error	121.2667	87	1.393870		

Table 3.12 Planned comparisons performed on the interaction between condition and problem type under the training group – neutral and inconsistent

	Sum of	Degr. of	Mean	F	p
M1	0.4167	1	0.416667	0.309388	0.579484
Error	117.1667	87	1.346743		

Table 3.12 ANOVA performed on the evidence evaluation ratings

	SS	Degr. of	MS	F	p
Intercept	40872.90	1	40872.90	1596.901	0.000000
Condition	697.65	2	348.83	13.629	0.000007
Error	2226.78	87	25.60		
Problem type	1432.45	2	716.23	74.534	0.000000
Problem type x condition	76.19	4	19.05	1.982	0.099267
Error	1672.02	174	9.61		

Table 3.13 Post-hoc analysis on condition

	Condition	control	instruction	training
1	control			
2	instruction	0.006050		
3	training	0.018459	0.000001	

Table 3.14 Post-hoc analysis on problem type

	PROBTYP	consistent	inconsistent	neutral
1	Consistent			
2	Inconsistent	0.000000		
3	neutral	0.009547	0.000000	

Table 3.15 ANOVA performed on the persuasiveness ratings

	SS	Degr. of	MS	F	p
Intercept	35984.03	1	35984.03	1478.090	0.000000
Condition	404.96	2	202.48	8.317	0.000495
Error	2118.01	87	24.34		
Problem type	1464.87	2	732.43	80.293	0.000000
Problem type x condition	61.91	4	15.48	1.697	0.152849
Error	1587.22	174	9.12		

Table 3.16 Post-hoc analysis on condition

	Condition	control	instruction	training
1	Control		0.004515	0.314285
2	Instruction	0.004515		0.000171
3	training	0.314285	0.000171	

Table 3.17 Post-hoc analysis on problem type

	Problem type	consistent	inconsistent	neutral
1	consistent			
2	inconsistent	0.000000		
3	neutral	0.000219	0.000000	

Table 3.18 ANOVA performed on Experiment Evaluation reasoning

	SS	Degr. of	MS	F	p
Intercept	1925.337	1	1925.337	405.4159	0.000000
Condition	12.496	2	6.248	1.3157	0.273580
Error	413.167	87	4.749		
Problem type	168.941	2	84.470	58.7444	0.000000
Problem type x condition	4.859	4	1.215	0.8448	0.498529
Error	250.200	174	1.438		

Table 3.19 Post-hoc analysis on problem type

	Problem type	consistent	inconsistent	neutral
1	Consistent		0.000000	0.000000
2	Inconsistent	0.000000		0.000004
3	neutral	0.000000	0.000004	

Table 3.20 ANOVA performed on evidence evaluation ratings

	SS	Degr. of	MS	F	p
Intercept	45916.45	1	45916.45	1360.042	0.000000
Condition	215.34	2	107.67	3.189	0.046065
Error	2937.21	87	33.76		
Problem type	1381.56	2	690.78	79.704	0.000000
Problem type x condition	64.41	4	16.10	1.858	0.119933
Error	1508.02	174	8.67		

Table 3.21 Post-hoc analysis on condition

	Condition	control	instruction	training
1	Control			
2	Instruction	0.031901		
3	training	0.030933	0.989794	

Table 3.22 Post-hoc analysis on problem type

	Problem type	consistent	inconsistent	neutral
1	Consistent			
2	Inconsistent	0.00		
3	neutral	0.00	0.007983	

Table 3.23 ANOVA performed on the argument validity ratings

	SS	Degr. of	MS	F	p
Intercept	40627.20	1	40627.20	1415.160	0.000000
Condition	273.16	2	136.58	4.757	0.010946
Error	2497.64	87	28.71		
Problem type	1250.02	2	625.01	76.935	0.000000
Problem type x condition	28.42	4	7.11	0.875	0.480360
Error	1413.56	174	8.12		

Table 3.24 Post-hoc analysis on condition

	Condition	control	instruction	training
1	Control			
2	Instruction	0.234812		
3	training	0.002939	0.065684	

Table 3.25 Post-hoc analysis on problem type

	Problem type	{1}	{2}	{3}
1	Consistent		0.000000	0.000000
2	Inconsistent	0.000000		0.000011
3	neutral	0.000000	0.000011	

Table 4.1 T-test for difference between obligation and permission problems pre-training

	Mean	Std.Dv.	N	Diff.	Std.Dv.	t	df	p
Obligation	0.491667	0.426770						
permission	0.625000	0.428092	60	-0.133333	0.366692	-2.81652	59	0.006592

Table 4.2 T-test for difference between obligation and arbitrary problems pre-training

	Mean	Std.Dv.	N	Diff.	Std.Dv.	t	df	p
Obligation	0.491667	0.426770						
arbitrary	0.100000	0.265215	60	0.391667	0.417739	7.262523	59	0.000000

Table 4.3 T-test for difference between permission and arbitrary problems pre-training

	Mean	Std.Dv.	N	Diff.	Std.Dv.	t	df	p
Permission	0.625000	0.428092						
arbitrary	0.100000	0.265215	60	0.525000	0.450282	9.031294	59	0.000000

Table 4.4 ANOVA (2 condition x 3 problem type x 2 prepost scores) on selection task responding

	SS	Degr. of	MS	F	p
Intercept	80.98767	1	80.98767	183.2963	0.000000
Training	0.14601	1	0.14601	0.3305	0.567616
Error	25.62674	58	0.44184		
Pre/post	1.70156	1	1.70156	8.8476	0.004272
Pre/post x training	0.23767	1	0.23767	1.2358	0.270863
Error	11.15451	58	0.19232		
Problem type	13.88993	2	6.94497	68.1082	0.000000
Problem x Training	1.17743	2	0.58872	5.7734	0.004071
Error	11.82847	116	0.10197		
Pre/post x problem type	0.28438	2	0.14219	2.4433	0.091336
Pre/post x problem type x training	0.52743	2	0.26372	4.5315	0.012737
Error	6.75069	116	0.05820		

Table 4.5 Follow-up analysis on the 3-way interaction under the obligation training – pre to post scores on the obligation problems

	Sum of	Degr. of	Mean	F	p
M1	1.066667	1	1.066667	6.649351	0.012479
Error	9.304167	58	0.160417		

Table 4.6 Follow-up analysis on the 3-way interaction under the obligation training – pre to post scores on the permission problems

	Sum of	Degr. of	Mean	F	p
M1	0.337500	1	0.337500	4.440454	0.039431
Error	4.408333	58	0.076006		

Table 4.7 Follow-up analysis on the 3-way interaction under the obligation training – pre to post scores on the arbitrary problems

	Sum of	Degr. of	Mean	F	p
M1	0.337500	1	0.337500	4.668820	0.034858
Error	4.192708	58	0.072288		

Table 4.8 Follow-up analysis on the 3-way interaction under the rules training – pre to post scores on the obligation problems

	Sum of	Degr. of	Mean	F	p
M1	0.004167	1	0.004167	0.025974	0.872524
Error	9.304167	58	0.160417		

Table 4.9 Follow-up analysis on the 3-way interaction under the rules training – pre to post scores on the permission problems

	Sum of	Degr. of	Mean	F	p
M1	0.004167	1	0.004167	0.054820	0.815703
Error	4.408333	58	0.076006		

Table 4.10 Follow-up analysis on the 3-way interaction under the rules training – pre to post scores on the arbitrary problems

	Sum of	Degr. of	Mean	F	p
M1	1.001042	1	1.001042	13.84795	0.000450
Error	4.192708	58	0.072288		

Table 4.11 ANOVA performed on changes in p only response patterns after training

	SS	Degr. of	MS	F	p
Intercept	116.2042	1	116.2042	32.90269	0.000000
TRAINING	11.7042	1	11.7042	3.31398	0.073853
Error	204.8417	58	3.5318		
PROBLEM TYPE	14.5042	1	14.5042	8.80497	0.004359
PROBLEM TYPE X TRAINING	5.7042	1	5.7042	3.46280	0.067835
Error	95.5417	58	1.6473		
PRE/POST	1.2042	1	1.2042	0.90186	0.346224
PRE/POST X TRAINING	5.1042	1	5.1042	3.82277	0.055385
Error	77.4417	58	1.3352		
PROBLEM TYPE X PRE/POST	0.3375	1	0.3375	0.71945	0.399812
PROBLEM TYPE X PREPOST X TRAINING	1.2042	1	1.2042	2.56692	0.114553
Error	27.2083	58	0.4691		

Table 4.12 ANOVA performed on changes in p/q response patterns after training

	SS	Degr. of	MS	F	p
Intercept	273.0667	1	273.0667	60.01086	0.000000
TRAINING	6.0167	1	6.0167	1.32226	0.254906
Error	263.9167	58	4.5503		
PROBLEM TYPE	81.6667	1	81.6667	38.41060	0.000000
PROBLEM TYPE X TRAINING	6.0167	1	6.0167	2.82984	0.097904
Error	123.3167	58	2.1261		
PRE/POST	8.0667	1	8.0667	7.80862	0.007036
PRE/POST X TRAINING	6.0167	1	6.0167	5.82420	0.018987
Error	59.9167	58	1.0330		
PROBLEM TYPE X PREPOST	4.2667	1	4.2667	7.51797	0.008110
PROBLEM TYPE X PREPOST X TRAINING	2.8167	1	2.8167	4.96304	0.029788
Error	32.9167	58	0.5675		

Table 5.1 ANOVA performed on problem types pre-training

	SS	Degr. of	MS	F	p
Intercept	36.85208	1	36.85208	181.1368	0.00
Error	24.21042	119	0.20345		
PROBLEM TYPE	24.27708	3	8.09236	107.3552	0.00
Error	26.91042	357	0.07538		

Table 5.2 ANOVA performed to examine difference between Cheng et al.'s and Klaczynski et al.'s training

	SS	Degr. of	MS	F	p
Intercept	73.17710	1	73.17710	397.3518	0.000000
{1}deontic/abstract training	0.07293	1	0.07293	0.3960	0.530391
{2}Klacz/Cheng training	1.99133	1	1.99133	10.8129	0.001335
D/A training x K/C training	2.18925	1	2.18925	11.8876	0.000788
Error	21.36279	116	0.18416		
{3}Pre/post	6.55279	1	6.55279	73.3316	0.000000
Prepost*D/Atraining	0.13057	1	0.13057	1.4612	0.229198
Prepost*K/Ctraining	2.12224	1	2.12224	23.7497	0.000004
Prepost*D/Atrain*K/Ctraining	0.53890	1	0.53890	6.0308	0.015541
Error	10.36557	116	0.08936		
{4}Problem type	18.83508	1	18.83508	222.8446	0.000000
Problem type x D/Atraining	0.86842	1	0.86842	10.2746	0.001743
Problem type x K/Ctraining	0.00071	1	0.00071	0.0084	0.927188
Problem type x D/Atrain x K/Ctrain	0.03474	1	0.03474	0.4110	0.522736
Error	9.80446	116	0.08452		
Prepost x Problem type	0.27953	1	0.27953	7.5671	0.006899
Prepost x Problem type x D/Atrain	1.24203	1	1.24203	33.6231	0.000000
Prepost x Problem type x K/Ctrain	0.04064	1	0.04064	1.1002	0.296412
3*4*1*2	0.27953	1	0.27953	7.5671	0.006899
Error	4.28501	116	0.03694		

Table 5.3 Planned comparison of pre to post training scores after Cheng et al.'s training

	Sum of	Degr. of	Mean	F	p
M1	0.60836	1	0.608362	6.808120	0.010270
Error	10.36557	116	0.089358		

Table 5.4 Planned comparison of pre to post training scores after Klaczynski et al.'s training

	Sum of	Degr. of	Mean	F	p
M1	8.06667	1	8.066667	90.27324	0.000000
Error	10.36557	116	0.089358		

Table 5.5 ANOVA performed on pre/post training scores for problem type under rules and obligation training

	SS	Degr. of	MS	F	p
Intercept	25.51276	1	25.51276	127.1319	0.000000
TRAINING	0.73151	1	0.73151	3.6452	0.061183
Error	11.63941	58	0.20068		
PREPOST SCORES	0.60836	1	0.60836	7.7262	0.007325
PREPOST X TRAINING	0.06947	1	0.06947	0.8823	0.351466
Error	4.56696	58	0.07874		
PROBLEM TYPE	9.30234	1	9.30234	95.2502	0.000000
PROBLEM X TRAINING	0.62526	1	0.62526	6.4023	0.014134
Error	5.66441	58	0.09766		
PREPOST X PROBLEM	0.05350	1	0.05350	1.6377	0.205732
PREPOST X PROBLEM X TRAINING	0.17156	1	0.17156	5.2516	0.025577
Error	1.89473	58	0.03267		

Table 5.6 Planned comparison on deontic problems pre to post scores under obligation training

	Sum of	Degr. of	Mean	F	p
M1	0.376042	1	0.376042	5.776000	0.019465
Error	3.776042	58	0.065104		

Table 5.7 Planned comparison on arbitrary problems pre to post scores under obligation training

	Sum of	Degr. of	Mean	F	p
M1	0.185185	1	0.185185	3.999310	0.050210
Error	2.685648	58	0.046304		

Table 5.8 Planned comparison on deontic problems pre to post scores under rules training

	Sum of	Degr. of	Mean	F	p
M1	0.004167	1	0.004167	0.064000	0.801176
Error	3.776042	58	0.065104		

Table 5.9 Planned comparison on arbitrary problems pre to post scores under rules training

	Sum of	Degr. of	Mean	F	p
M1	0.337500	1	0.337500	7.288743	0.009078
Error	2.685648	58	0.046304		

Table 5.10 ANOVA performed on pre/post training scores for problem type under permission and abstract training

	SS	Degr. of	MS	F	p
Intercept	49.65567	1	49.65567	296.1963	0.000000
TRAINING	1.53067	1	1.53067	9.1305	0.003737
Error	9.72338	58	0.16764		
PREPOST SCORES	8.06667	1	8.06667	80.6860	0.000000
PREPOST X TRAINING	0.60000	1	0.60000	6.0014	0.017334
Error	5.79861	58	0.09998		
PROBLEM TYPE	9.53345	1	9.53345	133.5589	0.000000
PROBLEM X TRAINING	0.27789	1	0.27789	3.8932	0.053255
Error	4.14005	58	0.07138		
PREPOST X PROBLEM	0.26667	1	0.26667	6.4707	0.013654
PREPOST X PROBLEM X TRAINING	1.35000	1	1.35000	32.7577	0.000000
Error	2.39028	58	0.04121		

Table 5.11 Planned comparison of pre to post scores on deontic problems under permission training

	Sum of	Degr. of	Mean	F	p
M1	1.837500	1	1.837500	22.55556	0.000014
Error	4.725000	58	0.081466		

Table 5.12 Planned comparison of pre to post scores on arbitrary problems under permission training

	Sum of	Degr. of	Mean	F	p
M1	0.504167	1	0.504167	8.441860	0.005183
Error	3.463889	58	0.059722		

Table 5.13 Planned comparison of pre to post scores on deontic problems under abstract training

	Sum of	Degr. of	Mean	F	p
M1	0.937500	1	0.937500	11.50794	0.001254
Error	4.725000	58	0.081466		

Table 5.14 Planned comparison of pre to post scores on arbitrary problems under abstract training

	Sum of	Degr. of	Mean	F	p
M1	7.004167	1	7.004167	117.2791	0.000000
Error	3.463889	58	0.059722		

Table 5.15 ANOVA performed on changes in p only response patterns under obligation and rules training

	SS	Degr. of	MS	F	p
Intercept	4.004167	1	4.004167	24.79755	0.000006
TRAINING	0.001852	1	0.001852	0.01147	0.915087
Error	9.365509	58	0.161474		
PROBLEM TYPE	0.266667	1	0.266667	7.81382	0.007019
PROBLEM X TRAINING	0.000463	1	0.000463	0.01357	0.907681
Error	1.979398	58	0.034128		
PRE/POST	0.016667	1	0.016667	0.20291	0.654065
PRE/POST X TRAINING	0.056019	1	0.056019	0.68199	0.412287
Error	4.764120	58	0.082140		
PROBLEM X PREPOST	0.004167	1	0.004167	0.34512	0.559167
PROB*PREPOST*TRAINING	0.007407	1	0.007407	0.61355	0.436640
Error	0.700231	58	0.012073		

Table 5.16 ANOVA performed on changes in p/q response patterns under obligation and rules training

	SS	Degr. of	MS	F	p
Intercept	31.90104	1	31.90104	112.7213	0.000000
TRAINING	0.00741	1	0.00741	0.0262	0.872039
Error	16.41447	58	0.28301		
PROBLEM TYPE	4.40104	1	4.40104	49.5328	0.000000
PROBLEM X TRAINING	0.18519	1	0.18519	2.0842	0.154208
Error	5.15336	58	0.08885		
PREPOST	0.53519	1	0.53519	5.1710	0.026685
PREPOST X TRAINING	0.11123	1	0.11123	1.0747	0.304195
Error	6.00289	58	0.10350		
PROBLEM X PREPOST	0.11852	1	0.11852	3.1165	0.082765
PROB*PREPOST*TRAINING	0.15845	1	0.15845	4.1666	0.045788
Error	2.20567	58	0.03803		

Table 5.17 ANOVA performed on changes in p only response patterns under permission and abstract training

	SS	Degr. of	MS	F	p
Intercept	2.140741	1	2.140741	19.47301	0.000045
TRAINING	0.535185	1	0.535185	4.86825	0.031329
Error	6.376157	58	0.109934		
PROBLEM TYPE	0.029630	1	0.029630	1.16200	0.285515
PROBLEM X TRAINING	0.001852	1	0.001852	0.07262	0.788509
Error	1.478935	58	0.025499		
PREPOST	0.255671	1	0.255671	4.02494	0.049503
PREPOST X TRAINING	0.126042	1	0.126042	1.98423	0.164284
Error	3.684259	58	0.063522		
PROBLEM X PREPOST	0.002894	1	0.002894	0.17295	0.679039
PROB*PREPOST*TRAINING	0.051042	1	0.051042	3.05081	0.085991
Error	0.970370	58	0.016731		

Table 5.18 ANOVA performed on changes in p/q response patterns under permission and abstract training

	SS	Degr. of	MS	F	p
Intercept	25.07989	1	25.07989	182.5356	0.000000
TRAINING	0.30697	1	0.30697	2.2342	0.140407
Error	7.96904	58	0.13740		
PROB	6.36461	1	6.36461	91.0942	0.000000
PROB*TRAINING	0.27225	1	0.27225	3.8966	0.053153
Error	4.05237	58	0.06987		
PREPOST	5.03151	1	5.03151	47.7477	0.000000
PREPOST*TRAINING	0.88614	1	0.88614	8.4092	0.005265
Error	6.11186	58	0.10538		
PROB*PREPOST	0.96901	1	0.96901	22.7523	0.000013
PROB*PREPOST*TRAINING	0.42364	1	0.42364	9.9470	0.002552
Error	2.47020	58	0.04259		

Table 5.19 T-test performed to investigate difference in ability between participants in Exps. 4 and 5 under obligation and rules training conditions

	t-value	df	p	F-ratio	p
Cognitive ability	-3.59750	118	0.000471	1.072548	0.788840

Table 5.20 ANOVA performed to investigate effectiveness of training between Exps. 4 and 5 for obligation and rules training

	SS	Degr. of	MS	F	p
Intercept	64.35013	1	64.35013	272.0709	0.000000
{1}TRAINING	0.36576	1	0.36576	1.5464	0.216174
{2}EXPERIMENT	0.77201	1	0.77201	3.2640	0.073408
TRAINING*EXPERIMENT	0.36576	1	0.36576	1.5464	0.216174
Error	27.43628	116	0.23652		
{3}PREPOST	1.94863	1	1.94863	19.7464	0.000020
PREPOST*TRAINING	0.10453	1	0.10453	1.0592	0.305530
PREPOST*EXPERIMENT	0.08578	1	0.08578	0.8692	0.353103
PREPOST*TRAINING*EXPERIMENT	0.00245	1	0.00245	0.0248	0.875198
Error	11.44716	116	0.09868		
{4}PROBLEM TYPE	19.30013	1	19.30013	186.5221	0.000000
PROBLEM*TRAINING	1.49076	1	1.49076	14.4071	0.000236
PROBLEM*EXPERIMENT	0.00638	1	0.00638	0.0617	0.804330
PROBLEM*TRAINING*EXPERIMENT	0.01055	1	0.01055	0.1019	0.750102
Error	12.00295	116	0.10347		
PREPOST*PROBLEM	0.19133	1	0.19133	5.1918	0.024523
PREPOST*PROBLEM*TRAINING	0.52779	1	0.52779	14.3216	0.000245
PREPOST*PROBLEM*EXPERIMENT	0.01217	1	0.01217	0.3302	0.566680
3*4*1*2	0.01981	1	0.01981	0.5374	0.464976
Error	4.27494	116	0.03685		

Table 6.1 ANOVA to investigate differences in responding dependent on problem type, training and order of task presentation

	SS	Degr. of	MS	F	p
Intercept	800.0000	1	800.0000	578.6416	0.000000
condition	10.8889	1	10.8889	7.8760	0.005735
order	3.2708	2	1.6354	1.1829	0.309476
condition*order	1.0486	2	0.5243	0.3792	0.685098
Error	190.7917	138	1.3825		
Problem type	539.0139	1	539.0139	402.3471	0.000000
Problem*condition	0.1250	1	0.1250	0.0933	0.760476
problem*order	0.6736	2	0.3368	0.2514	0.778060
prob*condition*order	1.3125	2	0.6562	0.4899	0.613775
Error	184.8750	138	1.3397		

Table 6.2 ANOVA to investigate effects of training on p only response patterns

	SS	Degr. of	MS	F	p
Intercept	36.1250	1	36.12500	32.31339	0.000000
condition	15.1250	1	15.12500	13.52913	0.000333
Error	158.7500	142	1.11796		
PROBLEM	8.6806	1	8.68056	13.80249	0.000291
PROBLEM*condition	4.0139	1	4.01389	6.38227	0.012623
Error	89.3056	142	0.62891		

Table 6.3 ANOVA to investigate effects of training on p/q response patterns

	SS	Degr. of	MS	F	p
Intercept	722.0000	1	722.0000	374.3460	0.000000
condition	1.1250	1	1.1250	0.5833	0.446291
Error	273.8750	142	1.9287		
PROBLEM	369.0139	1	369.0139	213.0200	0.000000
PROBLEM*condition	2.0000	1	2.0000	1.1545	0.284424
Error	245.9861	142	1.7323		