SOME EFFECTS OF STIMULUS VERIFICATION, INTERVIEW INSTRUCTIONS AND CONFIDENCE ON EYEWITNESS MEMORY.

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by

Mark Rhys Kebbell.

September 1995

ACKNOWLEDGEMENTS

Firstly, I would like to take this opportunity to express my thanks to my supervisor Graham Wagstaff, for his consistently excellent supervision throughout the duration of my studentship. Secondly, I would like to thank the many other people without whom I would never have been able to complete this PhD. These include; my mother and father, Alison Whybrow, Daryl 'O' Connor, Judith Covey, Liz Whelen, Ghulam Nabi, Pat Gwyer, Sue Thomas, Nicola Koyama, Merseyside Police and the Secretarial and Technical staff of the Psychology Department.

This PhD is dedicated to the memory of Mark Berry.

I would also like express thanks to the Economic and Social Research Council for funding this PhD studentship (Award No. R00429234159).

ABSTRACT

By the very nature of the way in which crimes are investigated, if little information is available from other sources, then greater efforts may be made to obtain information from eyewitnesses. This may include simple attempts to exhort the witness to try harder, or the use of 'special' memory facilitation procedures such as 'hypnosis' or the 'cognitive interview'. Thus, if considerable pressure is applied to an eyewitness to recall more information, it is likely to be in cases where very little of what the eyewitness recalls can be verified. However, these factors may make the eyewitness feel that there is considerable pressure to recall more information. If the eyewitness is aware that many, or indeed all of the answers that he/she provides cannot be verified, he/she may lower his/her criteria for report, confabulate, and express false confidence. In such cases where an eyewitness's testimony cannot be verified, investigating officers will tend to look to other measures, such as the confidence that an eyewitness expresses in his/her accounts, to determine accuracy. However, the experimental literature generally indicates that there is little, if any relationship between eyewitnesses' confidence and their accuracy and even this may be adversely influenced by interview procedures such as hypnosis.

These considerations give rise to three questions which were investigated in this thesis. These were as follows. 1) Does the knowledge that answers cannot be verified alter eyewitnesses' responses? 2) Is there a relationship between an eyewitness's confidence and his/her accuracy? 3) Are 1 and 2 mediated by interview instructions or memory facilitation techniques?

A preliminary series of six experiments was conducted to investigate the effects of verification on subject-eyewitnesses' memory reports on a photographic face-recognition task. Broadly speaking, subjects who were aware that their answers could not be verified did not appear to exaggerate their memory reports, even when procedures such as motivating instructions or hypnosis were used. However, there were a number of before/after instructions effects, indicating a general increase in subjects' estimates of their memory performance.

Two further experiments sought, more specifically, to investigate confidence-accuracy relationships. Subjects were shown film clips. This paradigm differed from previous work in this field because item difficulty was considered; i.e. questions were used that varied in difficulty from easy to hard. Using this procedure strong confidence-accuracy relationships were found, especially when subjects' 'absolutely certain' responses were considered.

A final major experiment sought to investigate the effects of hypnosis, a cognitive interview, and control procedures on accuracy and confidence, using a method which was a combination of those used previously. Once again verification had little effect on subjects' performance irrespective of interview condition. Confidence-accuracy relationships remained positive and high in each interview condition, although there was some evidence that hypnosis, but not the cognitive interview may increase confidence in both correct and incorrect information. It is suggested that these results may have practical implications for police interviewing.

PREFACE

Each year ten million people in the UK are interviewed by the police. Police detectives rate eyewitness information as being of central importance in the investigation of crime, and complete and accurate accounts from eyewitnesses are a major predictor of whether crimes will be successfully solved. Furthermore, erroneous eyewitness accounts are usually a contributory factor to the false conviction of innocent individuals. Consequently, eyewitness recall of criminal events has a considerable impact on the criminal justice system.

The actual part that eyewitnesses play in criminal proceedings varies greatly from case to case. For example, if a burglar is apprehended by the police as he runs away from a house that he has just burgled, admits the offence and has also left fingerprints at the crime scene, then there is already enough evidence to convict the burglar of the offence, so it would be unnecessary for the police to conduct in-depth interviews of eyewitnesses.

However, in other cases an eyewitness may be the only individual who witnessed a crime and may also be the only initial source of evidence. By the very nature of the way in which crimes are investigated, if little information is available from other sources, then greater efforts may be made to obtain information from an eyewitness. Thus, if considerable pressure is applied to an eyewitness to recall more information, it is likely to be in cases where very little of what the eyewitness recalls can be verified. This in itself may make the eyewitness feel that there is considerable pressure to recall more information. If the eyewitness is aware that many, or indeed all of the answers that he/she provides cannot be verified, he/she may lower his/her criteria for report, confabulate, and express false confidence.

The pressure on an eyewitness to produce information may be exacerbated by the use of instructions to the eyewitness to motivate him/her to produce more information, especially when used in the context of 'special' memory facilitation procedures such as 'hypnosis' or the 'cognitive interview'. These memory facilitation techniques may contain explicit and/or implicit suggestions that memory will be enhanced, creating a social situation which may place great pressure on eyewitnesses to recall extra information. Again, if the eyewitness is aware that many, or indeed all of the answers that he/she provides cannot be verified, he/she may lower his/her criteria for report, confabulate, and express false confidence.

In such cases where an eyewitness's testimony cannot be verified with physical forensic information or from other eyewitness accounts, investigating officers must rely on other measures, such as the confidence that an eyewitness expresses in his/her accounts, to determine accuracy. Thus, if an eyewitness says that he/she is 'absolutely certain' that an assailant had black hair then more emphasis will be placed on that information in a future investigation than if the eyewitness says that he/she is simply 'guessing' that the assailant's hair colour was black. It is plausible that eyewitnesses' confidence in information that they provide, both correct and incorrect, may be mediated by the knowledge that answers cannot be verified, by interview instructions and techniques, or by both.

These considerations give rise to three questions which are addressed in the present thesis: 1) Does the knowledge that answers cannot be verified alter eyewitnesses' responses? 2) Is there a relationship between an eyewitness's confidence and his/her accuracy? 3) Are 1 and 2 mediated by the interview instructions or memory facilitation techniques?

The thesis is divided into three parts: An introductory section, an empirical research section and a concluding section. A summary of each chapter of these three sections is outlined below.

<u>Part 1: Literature review and introduction of experimental variables</u>

The introduction presents a review of the literature relevant to the three research questions outlined above.

Chapter 1 is a general overview of the eyewitness testimony literature. The importance of eyewitness testimony and of improving, if possible, eyewitness memory are discussed. In addition, ways of predicting eyewitnesses' accuracy are considered. Particular consideration is given to how useful knowledge of the conditions in which an eyewitness witnessed an event is in predicting the likelihood that information provided by the eyewitness is accurate, and whether an eyewitness's own judgement of the likelihood that information he/she has provided is accurate would be more appropriate.

Chapter 2 concerns a review of the relationship between eyewitnesses' confidence and their accuracy. It is concluded that the confidence which eyewitnesses' express in information that they provide has a considerable influence on the way in which investigative and jury processes are conducted. However, the experimental literature generally indicates that there is little, if any relationship between eyewitnesses' confidence and their accuracy. Consideration is given to whether this finding may occur because researchers have paid insufficient attention to item difficulty.

Chapter 3 concerns the way in which 'standard' police interviews are conducted and problems that may occur. Criteria that need to be met in order to improve police interviewing are discussed.

Chapter 4 concerns the use of a 'cognitive interview' technique to enhance eyewitnesses' memory. The literature is reviewed and it is concluded that the cognitive interview appears to have the potential to enhance eyewitnesses' memory. However, it is noted that it has not been evaluated with the stringency that has been applied to other techniques such as hypnosis.

Chapter 5 concerns the use of 'hypnosis' as an aid to eyewitnesses' memory. A brief theoretical overview is provided of the area of hypnosis, and the implications of various conceptions of hypnosis for eyewitness investigative interviewing are discussed. The effects of hypnosis on eyewitness performance are considered, especially whether hypnosis may, in some eyewitness situations, lead to a lowering of a criteria for report and an increase in the confidence expressed in both correct and incorrect answers.

Chapter 6 concerns the issue of verification in eyewitness testimony. It is noted that if there is no way of verifying information then judicial procedures must rely solely on eyewitness information and the confidence of witnesses to judge accuracy. The implications of this are discussed.

Chapter 7 presents a general introduction to the experimental programme. It is concluded that three broad areas warrant further investigation. Firstly, does the knowledge that answers cannot be verified alter eyewitnesses' answers? Secondly, is there a relationship between eyewitnesses' confidence and their accuracy? And finally, are the former two areas influenced by interview instructions or techniques given to eyewitnesses?

Part 2: Empirical research

The empirical research section describes the experimental programme that was conducted.

Verification programme

The central focus of verification programme was to address the question 'does the knowledge that answers cannot be verified alter eyewitnesses' responses?'

Chapter 8 describes Experiment 1. In this, the performance of a group whose answers could be verified and a group whose answers could not be verified were compared on a face-recognition task testing memory for photographs.

Chapter 9 describes Experiment 2, which was similar to Experiment 1 except subjects were asked to repeat the face-recognition task after being given motivating instructions.

Chapter 10 describes Experiment 3, which was similar to Experiment 2 except subjects were given motivating instructions before undertaking the face-recognition task.

Chapter 11 describes Experiment 4, which was similar to Experiment 2 except subjects were asked to repeat the face-recognition-task after being given leading-motivating instructions.

Chapter 11 describes Experiment 5, which was similar to Experiment 3 with leading motivating instructions. However, in this experiment subjects were explicitly told that their answers could not be verified. Furthermore, to make the face recognition task more difficult none of the faces that they were required to identify were present.

Chapter 12 presents preliminary conclusions regarding the effects of verification.

Confidence-accuracy (C-A) programme

The main focus of the confidence-accuracy (C-A) programme was on the question 'is there a relationship between eyewitnesses' confidence and their accuracy?'

Chapter 13 describes Experiment 6, which was an investigation into item difficulty and the confidence-accuracy relationship. Subjects were shown a video film and asked a series of questions. The questions ranged in difficulty, half were easy, half were hard. Subjects were given two alternative answers to each question but were forced to make a choice.

Chapter 14 describes Experiment 7, which was similar to that of Experiment 6, with the exception that questions were changed to be more realistic. Specifically they were devised so as to fall into one of three categories, easy, medium or hard. Also, instead of choosing one of two alternatives, open-ended questions were used, although subjects were forced to make a choice.

Chapter 15 presents general conclusions concerning Experiments 7 and 8.

Interview instructions, verification and confidence-accuracy programme

The principal aim of the combined verification and C-A programme was to assess the extent to which verification and the C-A relationship might be influenced by hypnosis and cognitive interview techniques.

Chapter 16 describes Experiment 8. This pilot experiment was similar to Experiment 2, except hypnosis was used before a second face recognition questionnaire was administered.

Chapter 17 describes Experiment 9. This was a combination of the verification experiments and the confidence-accuracy experiments with three interview conditions: control, cognitive interview and hypnosis.

Part 3: General Discussion and Conclusions

In this section the results of the experimental work are discussed, conclusions are drawn, and suggestions for future work are given.

Chapter 18 presents discussion and conclusions with regard to the verification programme.

Chapter 19 presents discussion and conclusions with regard to confidence-accuracy relationships.

Chapter 20 concerns implications of the experimental programme for future work.

CONTENTS

TABLES

APPENDICES

PART 1

LITERATURE REVIEW AND INTRODUCTION OF EXPERIMENTAL VARIABLES

OVERVIEW

Part one presents a review of the literature relevant to the three research questions. 1) Does the knowledge that answers cannot be verified alter eyewitnesses' responses? 2) Is there a relationship between an eyewitness's confidence and their accuracy? And, 3) are 1 and 2 mediated by the interview instructions or memory facilitation techniques?

Chapter 1 is a general overview of the eyewitness testimony literature. The importance of eyewitness testimony and of improving, if possible, eyewitness memory are discussed. Ways of predicting eyewitnesses accuracy are considered. Chapter 2 concerns a review of the relationship between eyewitnesses' confidence and accuracy. The influence of the confidence which eyewitnesses' express in information that they provide has on the way in which investigative and jury processes are conducted are also considered.

Chapters 3 to 5 concern interview techniques. Chapter 3 concerns the way in which 'standard' police interviews are conducted and potential problems that may occur. Criteria to improve police interviewing are discussed. Chapter 4 concerns the use of a 'cognitive interview' technique to enhance eyewitnesses' memory. Chapter 5 concerns the utility of 'hypnosis' for enhancing eyewitness memory. A brief theoretical overview is provided that addresses the question 'what is hypnosis?'

Chapter 6 concerns the issue of verification in eyewitness testimony. It is noted that if there is no way of verifying information then judicial procedures must rely solely on eyewitness information.

Chapter 7 concerns a general introduction to the experimental programme and the research questions to be addressed.

CHAPTER 1

GENERAL OVERVIEW OF THE EYEWITNESS TESTIMONY LITERATURE

1.1 The significance of eyewitness testimony

The eyewitness interview is a major part of modern Police work; it is estimated that ten million eyewitnesses are interviewed each year in the UK (Merseyside Police, personal communication). When investigating criminal acts, the testimony of eyewitnesses is often of crucial importance. If eyewitnesses are unable to identify criminals or remember details of crimes, then perpetrators may go unpunished, while the recall of inaccurate information may mislead Police investigations, and is thought to be the most important factor in wrongful convictions (Huff & Rattner, 1988; Radkin, 1964). Wrongful convictions mean that individuals are punished for a crime they did not commit *and* guilty individuals go unpunished.

The importance of eyewitness testimony in Western criminal justice systems is illustrated by research conducted by Sanders (1986). When Sheriffs' deputies and detectives in New York were asked the question, "What is the central and most important feature of criminal investigations?", the majority of respondents replied "eyewitnesses". Also, the report of the Rand Corporation in 1975 found that the major predictor of whether a crime was solved was the completeness and accuracy of eyewitness accounts, especially when there are no other leads.

Because of these considerations, the field of eyewitnesses' testimony has received considerable attention from psychologists for nearly a century (e.g. Whipple, 1909), and especially over the past twenty years. Wells (1978) suggests that the role of this research should be:

> ... to generate scientific knowledge that will maximize the chances that a guilty defendant will be justly convicted while minimizing the chances that an innocent defendant will be mistakenly convicted (p.1546).

1.2 Overview of the experimental literature

There is now a large body of experimental work concerning factors that may influence eyewitness testimony. Often these factors are categorised according to their influence on acquisition, retention or retrieval (e.g. Hollin, 1989; Loftus, 1981). A brief summary of this research is given below.

Kassin, Ellsworth and Smith (1989) identified a number of factors in the area of eyewitness testimony research that they felt *may* have sufficient empirical support to be considered reliable enough for a psychologist to testify about in court. They surveyed 63 'experts' in the field of eyewitness testimony to ascertain how reliable the 'experts' felt these phenomena to be. These findings are included, where applicable, in the following account.

1.2.1 Acquisition

Acquisition concerns the factors which determine whether the information is attended to or encoded in the first place. Most obviously, the length of time for which an incident is witnessed appears to influence memory. Loftus (1972) and Hintzman (1976) showed that if the time that subjects could view a picture was increased their accuracy rate also increased (see also, Clifford & Richards, 1977). Eighty-five per cent of Kassin et al.'s experts felt that the statement 'the less time an eyewitness has to observe an event, the less well he or she will remember it', was sufficiently reliable to testify to in court.

Kuehn (1974) also found that the light conditions at a scene influenced accurate recall. Subjects recalled more accurate information when an incident took place in the daytime as opposed to in twilight conditions (see also, Yarmey, 1986).

The Yerkes-Dodson (1908) law has been applied to the relationship between stress and the eyewitnesses' ability to accurately recall crime details. It suggests that performance is related to stress in a curvilinear fashion; very low stress, produces poor performance, while a moderate amount of stress enhances recall. Excessive stress again decreases recall. This conclusion is supported by several studies. Peters (1988) recruited subjects from a health clinic, where subjects later received inoculations. During their visits subjects met either a nurse that injected them or a researcher who asked them questions for an equal amount of time. Later subjects were asked to describe and identify the nurse or the researcher. Peters found that subjects were better at identifying the researcher than the nurse. He concluded that this was due to the increased arousal caused by the fear of injection in that condition and thus, that if arousal increases much beyond normal, accuracy of memory suffers (see also, Clifford & Hollin, 1981; Clifford & Scott, 1978; Loftus & Burns, 1982). Kassin et al. found that 71 percent of experts felt that the statement 'very high levels of stress impair the accuracy of eyewitness testimony', was sufficiently reliable to testify to in court. Related to stress is the issue of violence. Kassin et al. found that 36 percent of experts felt that the statement 'eyewitnesses have more difficulty remembering violent than non violent events', was sufficiently reliable to testify to in court.

An issue related to stress and violence is that of weapon focus. Weapon focus is the concept that use of a weapon by a criminal will reduce the amount of accurate information that eyewitnesses can recall. It is hypothesised to occur because the weapon will draw the eyewitnesses' attention, to the exclusion of other information. Maass and Kohnken (1989) used a real-life simulation of an assault using a syringe as a weapon. They found a decrease in identification accuracy in the weapon group compared to controls (see also, Loftus, Loftus & Messo, 1987). Of Kassin et al.'s experts 57 per cent felt that the statement 'the presence of a weapon impairs an eyewitnesses' ability to accurately identify the perpetrator's face', was reliable enough to testify to in court.

A number of studies have shown that eyewitnesses' attitudes and expectations can affect their perceptions and thus their memory of an event. For example, Hastorf and Cantril (1954) asked opposing football fans to evaluate a football match between their two teams. The film was shown to both teams' fans, who were asked to count instances of inappropriate behaviour and fouls. In general each group reported many more fouls from their rivals than they did from their own team, their expectations influenced their reports. In another study, Peterson (1976) showed subjects a video film of a disturbance at a forum with Richard

Nixon. Subjects were given one of two sets of information before seeing the film. They were informed that the main characters on the tape were 'disruptive radicals' who intended to prevent the speaker from continuing, or alternatively that they were 'free-speech advocates' trying to ensure that both sides were heard. Subjects who expected to see angry radicals remembered more details consistent with this view and fewer details in opposition to this view. Fruzzetti et al. (1992) suggest that if five people witness an event there will be five different versions of what occurred. Eighty-seven percent of Kassin et al.'s experts felt that the statement 'an eyewitness's perception and memory for an event may be affected by his or her attitudes and expectations' was reliable enough to testify to in court.

Some experimental studies have shown a phenomena known as 'unconscious transference'. This, is the idea that an eyewitness may inadvertently confuse a criminal with an innocent bystander and so falsely identify the innocent bystander as the criminal. For example, Buckhout (1974) found that in response to a staged crime subject-eyewitnesses disproportionately identified bystanders to the crime as the criminal from perpetrator-absent line-ups. Given the statement 'eyewitnesses sometimes identify as a culprit someone they have seen in another situation or context', 85% of experts felt that this was reliable enough to testify to in court (Kassin et al., 1989).

Other factors, present at encoding have been shown to influence eyewitness memory, these include; race of target, race of witness; sex; crime seriousness; and whether information was central or peripheral to the event (Buckhout & Regan, 1988; Easterbrook, 1959; Leippe, Wells & Ostrom, 1978; Shaw & Skolnick, 1994).

1.2.2 Retention

Retention concerns the factors which influence how and whether information that is attended to is stored or retained. Common sense suggests that memory becomes less accurate over time (e.g. Ebbinghaus, 1913; Hollin, 1989). For example, Shepherd (1983) concluded that experimental results tended to suggest that 'lapse of time reduces the chance of errorless identification. If this is so, we should expect criminal cases in which there was a long delay between the offence and the identification parade to show more disagreement among witnesses and lower rates of identification than those in which the delay was much shorter' (p.177). Given the statement 'the less time an eyewitness has to observe an event, the less well he or she will remember it', 84% of experts felt that this was reliable enough to testify to in court (Kassin et al., 1989).

However, while delay before recall may decrease accuracy, post-event information before retrieval may have a greater adverse effect. Alper, Buckhout, Chern, Harwood and Slomovits (1976) showed subjects a staged incident. Subjects' memory was first tested with individual recall, followed by a group discussion and group memory for the incident. The group consensus descriptions were more complete than the individual descriptions but there was also more confabulation (see also, Hollin & Clifford, 1983; Meudel, Hitch & Kirby, 1992; Warnick & Sanders, 1980). The effects of hearing or reading descriptions and seeing photographs prior to making an identification has been investigated by a number of researchers. Loftus and Greene (1980) staged an incident in which witnesses viewed a central individual. Afterwards they read a script of the incident that included erroneous details concerning the target. When later asked to recall the target, some subjects incorporated the misleading information into their descriptions (see also, Davies, Shepherd & Ellis, 1979; Gorestein & Ellsworth, 1980). Given the statement 'Eyewitness testimony about an event often reflects not only what they actually saw but information they obtained later on', 87% of experts felt that this was reliable enough to testify to in court (Kassin et al., 1989).

1.2.3 Retrieval

At the retrieval stage the emphasis is on the witness's ability to retrieve information from memory. In many cases retrieval takes place under questioning (e.g. Fisher, Geiselman & Raymond, 1987; George, 1990; c.f. chapter three), consequently a number of studies have investigated the effects of different types of question on eyewitness memory. Loftus and Palmer (1974) showed subjects a filmed road accident. They were asked to estimate the speed of the cars when 'they ..... into each other'. Different groups of witnesses were given different words to fill in the blank. The words were 'contacted', 'hit', 'bumped' or 'smashed'. It was found that subjects' estimates of the vehicle's speed increased with the severity of the verbs; from 31.8 to 40.8 miles per hour (see also, Loftus, Altman & Geballe, 1975; Loftus & Zanni, 1975).

Questioning may be a source of misleading information. Loftus (1975) showed subjects a film of a car on a country road. When questioned half the subjects had to judge the speed of the car 'as it passed the barn' a control group were asked to estimate the speed of the car without reference to the fictitious barn. One week later subjects were asked if they had seen a barn. Significantly more subjects that had been given misleading information stated that they could remember a barn than the control group (see also, Loftus, 1975, Miller & Loftus, 1976). Ninety-seven per cent of Kassin et al.'s experts felt that the statement 'An eyewitness's testimony about an event can be affected by how the questions put to that witness are worded', was sufficiently reliable to testify to in court.

Of importance is the witnesses' ability to assess the accuracy of that information they provide (i.e. the confidence they express in the accuracy of information they provide). The majority of studies suggest that there is either no relationship or only a small positive relationship between eyewitnesses' confidence and their accuracy (e.g. Deffenbacher, Brown & Sturgil, 1978; Fruzzetti, Toland, Teller & Loftus, 1992; Lindsay & Wells, 1980; Smith, Kassin & Ellsworth, 1989; Wells & Leippe, 1981; c.f. chapter two). Indeed, when Kassin et al.'s experts were asked to evaluate the statement 'an eyewitness's confidence is not a good predictor of his or her identification accuracy', 87 per cent stated that this statement was reliable enough to testify to in court.

More recently attention has been focused on methods of enhancing eyewitness memory retrieval though interview procedures. Probably the most popular are the cognitive interview (c.f. chapter four, e.g. Fisher, Geiselman & Amador, 1989; Fisher, Geiselman & Raymond, 1987; Fisher, Geiselman, Raymond, Jurkevich & Warhaftig, 1987; Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissan & Prosk, 1984; Geiselman, Fisher, MacKinnon & Holland, 1985; Geiselman, Fisher, MacKinnon & Holland, 1986; Geiselman & Padilla, 1988) and hypnosis (e.g. Sheehan, 1988; Smith, 1983; Wagstaff, 1995; c.f. chapter five). Although Kassin et al. did not ask their experts about the cognitive interview, they did ask two questions about the influence that hypnosis may have on eyewitnesses. In response to the question 'hypnosis *does not* facilitate the retrieval of an eyewitness's memory', 52% of Kassin et al.'s experts felt that this was reliable enough to testify to in court. Sixty-nine per cent of these experts felt that the statement 'hypnosis increases the suggestibility to leading and misleading questions' was reliable enough to testify to in court.

1.3 The usefulness of eyewitness testimony research

Despite the resources that have been devoted to investigations of factors that may effect eyewitnesses' testimony, the usefulness of this work has been called into question (e.g. Egeth, 1993; Egeth & McCloskey, 1984; Elliott, 1993; Konecni & Ebbeson, 1986; McCloskey, Egeth & McKenna, 1986; Wells, 1978). Criticisms focus on the reliability of the eyewitness testimony research and the usefulness of this research and the relevant literature in making *specific* estimates of an eyewitness's accuracy in a given situation.

Indeed, many of the findings that Kassin et al.'s experts were prepared to testify to in court appear to have questionable reliability. For example, the view that high levels of stress necessarily adversely effects eyewitnesses' recall seems debateable. Yuille and Cutshall (1986) considered the memory of eyewitnesses for a real shootout that involved a robber that was fatally shot and a store owner who was seriously wounded. Most of the witnesses were severely upset and stressed. But, those witnesses that said they were most stressed were actually *more* accurate than the less stressed witnesses. This is presumably because they had a more central view of the incident. But Elliott (1993) asks, would Kassin's experts have been able to predict this before the fact rather than being able to explain it afterwards?

Such enhancements of performance with high arousal have also been reported by Tooley, Brigham, Maas and Bothwell (1987) who found a positive association of face recognition with induced arousal using noise and threat of electric shock, but whether these effects would have become negative with the threat levels associated with typical crimes is unclear. Other studies have found no relationship between arousal and recall. Thus Christianson and Hubbinette (1993) found no effect of reported fear on eyewitnesses' memory of real armed robberies. Nightingale, Kebbell, Thomas and Wagstaff (1995) also evaluated the role of fear on the memory performance using first time parachutists, who were tested for their memory for the aircraft from which they had jumped. They found no decrease in performance in subjects that reported high levels of fear compared with subjects reporting low levels of fear. Thus it would appear that stress may improve, make no difference or adversely effect eyewitness performance depending on conditions which, as yet, are not clearly defined. One possible moderating factor is the centrality of details to be recalled.

Similar problems have occurred with other, supposedly well-established phenomena in the field of eyewitness testimony such as weapon-focus. Elliott (1993) reviews 10 experiments that have been conducted to investigate weapon-focus. Six of these studies were conducted by Penrod (Cutler & Penrod, 1988; Cutler, Penrod & Martens, 1987a, 1987b; Cutler, Penrod, O'Rourke & Martins, 1986; O'Rourke & Martins, 1986; O'Rourke, Penrod, Cutler & Stuve, 1989). Of these six studies, two effects were significant, four are not, and one was a trend in the direction against the hypothesis. Four other studies have shown a small effect (average variance 3 - 4%), three of which used potentially less life-like slide presentations (Kramer, Buckhout & Eugenio, 1990; Loftus, Loftus & Messo, 1987; Tooley et al., 1987) and that of Maass and Kohnken (1989). Furthermore, Elliott suggests that if an experiment does not yield a positive result then it is often not published, so it is possible that there maybe other studies that did not find weapon-focus effects that are what Elliott calls 'sitting in file drawers'. The above studies allow no definitive conclusions to be drawn concerning weapon-focus effects. It appears that weapon-focus does sometimes occur, but not always, and it is not clear why it sometimes does and sometimes does not occur.

Similar criticisms have been applied to the reliability of other effects in the eyewitness testimony literature, for example unconscious transference (e.g. Read, Tollestrup, Hammersley, McFadzen & Christensen, 1990) decreases in memory performance over time (e.g. Yuille & Cutshall, 1986), the adverse effect of erroneous post-event information (e.g. Yuille & Cutshall, 1986), and misleading question effects (e.g. McEwan & Yuille, 1981). Research on confidence-accuracy, and improving memory, will be considered in detail shortly.

Attempts have been made to answer these criticisms (see for example, Kassin, Ellsworth & Smith, 1994), but even if these effects were reliably reported in the experimental literature, it seems that such information could not predict in a given case whether a piece of an eyewitness's testimony was correct or incorrect. All that could be provided would be probabilistic judgements of accuracy.
Furthermore, in real situations several factors are likely to occur at the same time. For example, how likely is a witness to correctly identify an armed robber that he/she saw for a brief period of time in good light, armed with a pistol, after a period of six months? It is possible that these factors may interact with one

another. Wells (1978) notes that it is simply not possible to simply assess the likelihood of an eyewitness being accurate by 'plugging in' relevant factors. He gives the example of having a checklist such as:

> What is the victim's race? What is the defendant's race? How attractive was the defendant? What is the witness's sex? How old is the defendant? How severe was the crime? What was the witness's perceptual set? Visual context? Exposure time? And so on (p.1551).

Could such information be used to determine how likely an individual's memory is to be accurate? He states that such an approach is futile because of its complexity. To systematically evaluate the relationships between a variety of eyewitness variables requires an experimental design with a number of cells. In order to evaluate the relationship between 20 variables that could potentially interact with one another, a design would be required with 1,048,576 cells. If 10 subjects were assigned to each cell this would require over 10 million subjects.

In sum, while research in the field of eyewitness testimony has identified a number of factors that might influence eyewitness accuracy, such studies have been criticised for being insufficiently reliable, moreover, it has been argued that the factors are too complex for effects to be confidently predicted.

1.4 The importance of Confidence-Accuracy relationships

Although researchers have placed much emphasis on the factors discussed above, possibly the most important area of eyewitness testimony is that of the relationship between confidence and accuracy. Indeed, one could argue that this has had a profound effect on work that has been conducted in this field.

It is commonly assumed by researchers that there is no, or little positive, relationship between eyewitnesses' confidence and their accuracy (e.g. Deffenbacher, Brown & Sturgil, 1978; Fruzzetti, Toland, Teller & Loftus, 1992; Lindsay & Wells, 1980; Smith, Kassin & Ellsworth, 1989; Wells & Leippe, 1981; c.f. chapter two). Given this it is plausible that the principal reason for many of the various investigations of factors that may influence eyewitness accuracy may have stemmed from the supposition that eyewitnesses cannot accurately assess their own accuracy. If there were a positive relationship between eyewitnesses' confidence and their accuracy then it would be unnecessary to investigate the variety of other factors that might affect accuracy; all one would have to do would be to ask the eyewitness how confident he/she is in the accuracy of information he/she is providing.

Earlier an example was given that concerned the ability of a witness to correctly identify an armed robber that he/she saw for a brief period of time in good light, armed with a pistol, after a period of six months. If one could rely on confidence as a predictor of accuracy, then if the witness was 'absolutely certain' that he/she had correctly identified the robber, then that identification would be very likely to be correct and one would not have to worry about potential influences of a weapon, retention interval, etc. If the witness was 'not at all sure' that his/her identification was correct one would be aware that the identification may be incorrect and so should be treated with caution.

1.5 The importance of interview technique

Much of the research on eyewitness testimony has tended to emphasise the negative aspects of eyewitness performance, and to specify the circumstances in which eyewitnesses are most likely to produce errors (see for example, Bull & Clifford, 1979; Loftus, 1979). The findings of such research have led researchers such as Fisher, Geiselman and Raymond (1987) to conclude that eyewitness performance is 'incomplete, unreliable, partially constructed and malleable during the questioning procedure', and add that 'because of the potential inaccuracy in eyewitness reports, strict reliance on eyewitness identification may often lead to false convictions' (p. 401).

However, such research often focuses on variables that are not under the control of the Police. Wells (1978) terms these variables 'estimator' variables because in real crimes one can only *estimate* how much of an effect they will have. A more fruitful and forensically useful area for research might concern variables that are under the control of the Police and the courts. Wells calls these variables 'system' variables as they are part of the criminal justice system. For example, if the interval between witnessing an event and being interviewed influences accuracy (Shepard, 1967), it may be advisable to interview eyewitnesses for as soon as is reasonably practicable.

The central difference between system and estimator variables is that once a crime has been witnessed Police and courts are powerless to alter what the witness saw; however, it is possible to change system variables such as reducing the amount of time between when an eyewitness witnessed an event to when they were interviewed. Perhaps, therefore, it is appropriate to focus research on areas such as interview techniques, and particularly those which may enhance confidence-accuracy relationships, as such techniques can potentially be altered in the field.

CHAPTER 2

CONFIDENCE-ACCURACY RELATIONSHIPS IN EYEWITNESS TESTIMONY

2.1 Why is the relationship between eyewitness confidence and accuracy important?

The confidence eyewitnesses express in information has an important influence on both the investigative process and the credence which jurors give to eyewitness testimony. If eyewitnesses express certainty that their answers are correct, their responses are more likely to be perceived as correct (e.g. Brigham & Wolfskeil, 1983; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Stuve, 1988; Cutler, Penrod & Thomas, 1988; Fox & Walters, 1986; Leippe, Manion & Romanczyk, 1992; Lindsay, 1994; Lindsay, Wells & O'Connor, 1989; Lindsay Wells & Rumpel, 1981; Wells, Ferguson & Lindsay, 1981; Wells, Lindsay & Ferguson, 1979). Furthermore, surveys of the lay-public in the United States (Brigham & Bothwell, 1983; Deffenbacher & Loftus, 1982), Germany (Sporer, 1983), Canada (Yarmey & Jones, 1983), Australia (McConkey & Roche, 1989) and England (Noon & Hollin, 1987) reveal that there is a substantial, cross-cultural belief that confidence predicts accuracy.

These perceptions are in agreement with the 'cognitive' literature that indicates that there is a moderate yet robust, positive C-A relationship. This relationship exists over different populations, tasks and experimental materials (e.g., Blake, 1973; Hart, 1965, 1966, 1967; Nelson, Gerler & Narens, 1984; Perfect, Watson & Wagstaff, 1993; Schacter, 1983; Stephenson, 1984; Stephenson, Brandstatter & Wagner, 1983; Stephenson, Clarke & Wade, 1986; and for a review, Nelson, 1988).

However, although it might appear intuitively obvious that there is a strong positive relationship between *eyewitnesses'* confidence and their accuracy, much research in this latter area appears to contradict this assumption.

2.2 The empirical literature with respect to eyewitness C-A relationships

2.2.1 C-A relationships for identifications

Reviews that have considered the C-A relationship for identifications suggest that although the relationship varies greatly from study to study there is generally either no relationship, or only a small positive relationship between eyewitnesses' confidence and their accuracy (e.g. Bothwell, Deffenbacher, & Brigham, 1987; Deffenbacher, 1980; Fruzzetti, Tolland, Teller & Loftus, 1992; Penrod, Loftus & Winkler, 1982; Wells, 1993; Wells & Murray, 1984).

Kassin, Ellsworth and Smith (1989) also asked their 63 experts to evaluate the statement 'an eyewitness's confidence is not a good predictor of his or her identification accuracy'. Forty-four per cent of the respondents found the research supporting this statement to be 'very reliable', 29% found it to be 'generally reliable', 14% agreed that the research 'tends to favour' that conclusion, 10% felt that the work was inconclusive, 2% found no support for the statement and none thought that the reverse was true, i.e. that confidence was a good predictor of identification accuracy. In addition, 52% thought that the finding was reliable enough to testify in court about, 83% stated that they would testify about it, and 37% reported that they had testified about it. Indeed, in Kassin et al.'s survey, that the relationship between confidence and accuracy was weak with respect to identifications was found to be the fourth most reliable phenomenon that Kassin et al. considered in the eyewitness literature.

Attempts have been made to explain these apparently counter-intuitive findings (for a review see, Sheehan, 1988). Deffenbacher (1980) highlights differences in C-A relationships according to the conditions during encoding, storage and retrieval. He claims that one finds zero or even negative C-A relationships with conditions or test situations that are likely to produce errors or mistakes, but when low errors are likely, the relationship may be positive and quite strong. The hypothesis that Deffenbacher uses to explain this variability is expressed by him in terms of the degree of 'optimality' of the information-processing conditions present at encoding, storing and retrieval. Under ideal circumstances witnesses will accurately gauge the adequacy of their memory performance in their memory reports while under conditions that are not optimal, the two will covary less reliably, opening the way for other variables, such as personality or interview procedure, to determine the degree of confidence. While Leippe (1980) recognises this essential variability of the C-A relationship, he considers the optimality hypothesis in terms that also stress the integrative nature of subjects' memories and cognitions and the social factors that may influence confidence independently of accuracy.

According to Leippe, two characteristics of human memory and cognition-their integrative nature and their unconscious operation, may influence memory accuracy and confidence to move in different directions, especially when the test situations are highly influential, such as when leading questions are used. Leippe argues that as the operation of processes that alter or transform memory become more extensive, C-A relationships become smaller. Subjects have a feeling about the strength of their representations in memory, but it is unlikely that they will be conscious of the transformations that may have affected these representations during stages of encoding, storage and retrieval. The essential argument is that if we are not conscious of whether or to what extent internally produced alterations of our memories have taken place, then it is likely that we will be poor judges of the accuracy of our recollections if these alterations have occurred.

However, Sporer, Penrod, Read and Cutler (in press) have found considerably higher C-A relationships than have been previously reported by using choice as a moderator variable (see also, Brigham, 1988; Sporer, 1992, 1993, 1994). They categorise subjects as either choosers or non-choosers. Choosers are subjects that positively identify someone from a lineup, while non-choosers are subjects that reject the lineup. They argue that this is an appropriate paradigm, as there are different forensic outcomes for choosers and non-choosers. Choosers are likely to influence the legal system through identifications that they make, whether they are accurate and identify the target individual or if they are inaccurate but identify the individual that the police suspect of a crime. Non-choosers are less likely to have an impact within the legal system either because they would be seen as unreliable (particularly if it was determined or believed that the suspect had been missed) or because a non identification would not lead to criminal charges (see also, Malpass & Devine, 1981; Wells & Murray, 1984).

To investigate the influence of choice as a moderator variable for C-A relationships Sporer et al. conducted a meta-analysis of 30 studies in which target-present and target-absent lineups were presented to subjects. The total number of subjects in all experiments of 4036. For choosers the C-A correlation was reliable and consistently higher (r=.41), than for non-choosers (r=.12). Furthermore, mean confidence levels for correct choosers were higher than those for incorrect choosers in each of the 30 studies.

Thus, although there is a consensus amongst 'experts' that there is little or no relationship between confidence and accuracy, the relationship may be high in optimal conditions and with a different method of analysis may be greater than has been previously reported. However, it is important to note that while the overwhelming majority of work has been conducted concerns C-A relationships with respect to identifications, in typical eyewitness situations, more emphasis is likely to be placed on the confidence that eyewitnesses express in recall-information that they provide, for example, descriptions of a criminal. Indeed, eyewitnesses are very rarely asked to undertake line-up identifications, and because identification lineups conducted in line with the UK Codes of Practice are expensive in terms of Police resources they are becoming less common (Merseyside Police, personal communication).

2.2.2 C-A relationships for recall

Very little research has investigated C-A relationships with respect to information that eyewitnesses provide other than for identifications, although many have generalised findings from C-A for identification accuracy, possibly inappropriately, to include eyewitness C-A relationships for recall (e.g., Canter, 1994; Fisher & Geiselman, 1992).

Smith, Kassin and Ellsworth (1989) investigated C-A relationships for subjects' recall and attempted to indicate why there was little or no relationship between C-A outlined previously. Smith et al. (1989) suggested that researchers have concentrated on the C-A relationships 'between-subjects', comparing the accuracy of confident witnesses to less confident witnesses, rather than the relationship within subjects' own statements. In the latter case, an eyewitness may say that he/she is absolutely certain of some things but is not at all certain of others. To assess within-subject and between-subject confidence accuracy relationships Smith et al. (1989) showed subjects a slide presentation followed by a number of two-alternative forced choice questions. They were then required to rate their confidence in each answer on a ten-point scale. The average between-subjects and within-subjects C-A correlations were comparatively low, $r=.14$ and $r=.17$ for between and within subjects measures respectively. Smith et al. drew the following conclusion:

> Confidence is not a good predictor of accuracy. Common sense and the Supreme Court notwithstanding, confidence is not a useful indicator of the accuracy of a particular witness or of the accuracy of particular statements made by the same witness. The present data indicate that relying on confidence to assess the credibility of witness's statements may be dangerously misleading. Probably

evidence may be ignored because it is not confidently asserted and errors believed because the witness is certain (p.358).

However, Perfect, Watson and Wagstaff (1993) note that Smith et al. assessed memory with a forced two-choice recognition procedure that gave a hit rate of only 63%. As a large number of these hits (37%) would have occurred by chance, they suggest that this high guessing rate may be in part responsible for the low correlations found. Perfect et al. therefore attempted to reduce the chances of producing correct answers by guessing, by using a five-alternative forced-choice question format. They found an higher overall correlation than Smith et al. for the between-subjects analysis (Goodman-Kruskal Gamma =.49), but no correlation for the within-subjects analysis (Gamma=-.03).

Perfect et al. explain the discrepancy between findings concerning eyewitness-recall C-A relationships and that of the cognitive literature in a variety of ways. They highlight the differences between laboratory-based work and more forensically motivated work. Eyewitness memory is based on an event witnessed only once, under what is often non-optimal conditions that are often incidentally learned and may have a strong emotional component. Specific recall of details from episodic memory is required. All these aspects differ from the memory requirements of the typical feeling of knowing study, which evokes no emotional involvement and which tests memory for relatively familiar material and is more likely to rely on semantic memory. As well as differences in the types of encoding that takes place there are likely to be differences in the opportunities for confidence calibrations afforded by semantic and episodic memory. For semantic memory, they suggest, one has many opportunities to self-test one's ability at retrieving particular facts. Moreover, there are specified and agreed upon answers to general knowledge questions so one not only can gauge how appropriate one's confidence in an answer should be but also can determine one's relative performance with others in the same situation. For eyewitness memory there is no way of knowing that a specific item retrieved is correct because the event cannot be revisited. There is no agreed upon answer to calibrate one's performance against, nor can one know whether one's ability to recall an event is better or worse than anyone else because no two witnesses are likely to have seen the same event from exactly the same perspective.

However, the hypothesis, that eyewitness C-A relationships are poor because eyewitnesses are rarely able to check their answers and thus calibrate themselves, ignores the possibility that there may be many occasions in which an individual relates an event to a third party in the presence of another individual who was also present at that event (Luus & Wells, 1994). In such situations that individual is able to receive feedback if inaccuracies are detected by the other witness this may help calibrate that individual's future accounts.

Scogin, Calhoon and D'Errico (1994) investigated the C-A relationship for subjects in three age cohorts, young, medium and old. Subjects were presented with a short film, then they were given a three-alternative forced-choice questionnaire. Subjects were asked to rate their confidence in their answers on a scale from 1 (not at all confident) to 10 (perfectly confident). A small, but significant correlation, was obtained between mean confidence rating and percentage of questions answered correctly for all three groups combined ($r=.28$). Furthermore, when each persons' mean confidence ratings in correct answers were compared to confidence in incorrect answers, confidence was significantly greater for correct answers. Unfortunately, within-subjects C-A correlations were not calculated, though, clearly this study does show some positive relationship between C-A.

Thus, the experimental literature tends to suggest that there is not a strong relationship between eyewitness confidence and accuracy for recall information, even if 'within-subjects' relationships are considered.

2.2.3 Item difficulty

However, one factor that has yet to be systematically investigated is item difficulty. Typically in work in this area, researchers attempt to select items so as to avoid floor and ceiling effects; i.e. they try to avoid items that are either very easy or very hard to remember. But in real-life forensic investigations some questions that eyewitnesses are asked may clearly be easier to answer than others. For example, gender is very likely to be answered accurately, whereas eyecolour is less likely to be answered accurately (Christianson & Hubinette, 1993). Given this, it seems reasonable to propose that most eyewitnesses will be likely to be very confident that their identification of an individual's sex, but considerably less confident about their report of eye-colour. This is not a minor point; knowing the gender of the suspect is of considerable use in a forensic investigation, immediately eliminating half of the population from future enquiries. It may be the case, therefore, that previous researchers may have chosen unrealistic and overly homogeneous pools of items, thus reducing the variance necessary for high correlations.

Another potentially important factor is that of the relationship between 'absolutely certain' responses and accuracy. This effect may be precluded when 'easy' items are excluded, and, as it is not necessarily related to correlation size, it may often be missed in correlational analysis as incorrect responses are required in order to calculate C-A relationships (Gruneberg & Sykes, 1993). However, regardless of overall C-A accuracy, it could be the case that the relationship between these 'absolutely certain' responses and accuracy remains high.

2.3 Conclusions

In sum, there is a consensus of opinion amongst experts in eyewitness testimony that there is little relationship between confidence and accuracy where identification accuracy is concerned, although different experimental procedures and methods of calculating this relationship may increase this relationship. When recall information is considered, although there has been limited work in this area, C-A relationships do not appear to be large. However, further research into this area particularly with respect to item difficulty might be fruitful.

CHAPTER 3

THE STANDARD POLICE INTERVIEW

Given the wealth of evidence apparently indicating the unreliability of eyewitness reports psychologists have for some time been attempting to develop interviewing techniques for enhancing the volume and accuracy of information given by eyewitnesses (see for example, Malpass & Devine, 1981; Wagstaff, 1982). However, in the present thesis attention will be paid to the two most commonly applied psychological techniques, the cognitive interview and hypnosis.

As pointed out at the start of this thesis, one important question concerns how C-A relationships might be affected by interview techniques designed to facilitate memory. In order to fully address this question, it is first useful to describe the nature and aims of, and problems associated with, standard police interview procedures.

3.1 Why do the police interview eyewitnesses?

The UK Police Central Planning and Training Unit's (1992) booklet 'Investigative Interviewing: A Guide to Interviewing' gives three answers in response to the question 'why do I need to interview this person?' These are, 1) to find out what a witness or victim has seen or heard; 2) to obtain evidence to prove or disprove a suspect's involvement in an allegation; and 3), where evidence is already available, to interview further witnesses and test or confirm such previously obtained accounts (p.2). Thus, the police interview has a clear function; to collect evidence that can be used, if necessary, in future legal proceedings.

The way in which the Police conduct an investigation involves a number of stages and may involve several different methods of information gathering depending on the crime. Gudjonsson (1992) notes:

> Broadly speaking, most crimes are solved by the use of one or more of the following sources of information: 1) There may be witnesses to the crime and they need to be interviewed and possibly give evidence in court in due course. Victims and police officers are also potential witnesses. An identification parade may be set up if the police have a potential suspect. 2) Information may be supplied by informants, whose motivation to talk may include financial considerations, revenge or moral considerations. 3) Criminal suspects may give information to the police during interviewing, including self-incriminating admissions or confessions. 4) Forensic science techniques may provide the police with tangible evidence. This includes the work of the pathologist, the fingerprint expert, the forensic scientist, and the scene-of-crime officer (p.6).

The relative importance of each of these sources of information will vary from crime to crime, and will affect the way in which eyewitnesses are interviewed. For example, if a burglar is apprehended by police officers as he runs away from a house that he has just burgled, admits the offence and has also left fingerprints at the crime scene, then there is already enough evidence to convict the burglar of the offence, so it would be unnecessary for the police to conduct in-depth interviews with eyewitnesses.

However, in other cases an eyewitness may be the only individual that witnessed a crime and may also be the only initial source of evidence. In such cases if the crime is serious and if little information is available from other sources, then greater efforts may be made to obtain as much information as possible from an eyewitness. In these situations it is particulary important for the eyewitness to produce as much information as possible.

3.2 How do the police interview?

While the way in which suspects are interviewed has received considerable attention (e.g. see Baldwin, 1993; Gudjonsson, 1992), little formal attention has been paid to the methods that the police use to interview non-hostile eyewitnesses. However, Fisher and his colleagues conducted an analysis of real life Police interviews in the USA with the intention of taking their research 'Out of the laboratory and into the field where actual crime interviews are conducted by Police' (Fisher, Geiselman & Raymond, 1987, p.177). They concluded that there is considerable scope for improving interview techniques without recourse to complicated theoretical constructs. This work will be considered in some detail here.

Fisher, Geiselman and Raymond (1987) examined interviews conducted by Police officers in the state of Florida in the United States of America. Eleven tape recorded interviews, conducted by experienced detectives were analyzed. These interviews covered a range of crimes, for example, crimes committed with or without a weapon, with one or more suspects, in the day or evening, at the eyewitness' home, in a street or at the eyewitness' place of work. A possible limitation of this study is that so few tapes were used, nevertheless, an analysis of the interviews indicated the following common problems.

1.Interruption of Eyewitnesses' responses.

Fisher et al. found the major problem with the interviews was frequent interruption of the eyewitnesses' responses by the interviewing Police officer. After introducing themselves, all of the interviewers asked the eyewitness to tell them what had happened. However, during this free recall the eyewitness would be interrupted frequently. In the interviews there were, on average, three open-ended questions requiring an extended answer. During the responses to the open-ended questions the interviewer interrupted the eyewitness on average eleven times. In the typical interview the eyewitness was interrupted only 7.5 seconds after they had begun to reply.

They argue that these interruptions cause two main problems. Firstly, they break the concentration of the eyewitnesses when they are trying to retrieve information. The retrieval of information from one's memory is a difficult process at the best of times, but if the eyewitness' concentration is broken by an interviewer's question, then the eyewitness must switch attention from trying to recall information, to the interviewer's question, then back to their memory in order to answer the question. This makes the task much more difficult. Such constant shifting of attention prevents optimal recall of the event. This is particularly unfortunate as free recall typically produces very accurate recall. Further, the increased difficulty of trying to recall information despite constant interruptions may stop the eyewitness from trying so hard to recall information.

The second drawback of interruption is that after eyewitnesses have been interrupted several times they begin to expect to be interrupted throughout the interview. This leads the eyewitnesses to tailor their responses to fit the interview format. As the eyewitnesses expect to have only a short period to respond, they shorten their responses accordingly. Any response which is shortened will not produce as much information, and may exclude information which may be important to the forensic investigation.

Interruptions did not only take the form of questions from the interviewing officers; for example some eyewitnesses were interrupted by the Police officer's radio, and others by someone walking into the interview room - problems which could easily be avoided by turning the radio off or placing a 'do not disturb' sign on the door.

2. Excessive use of question-answer format.

Closely related to the problem of frequent interruption is the excessive use of a question-answer format. Fisher et al. categorised questions as either 'open-ended' questions, where eyewitnesses were required to give a complex response, such as "can you describe the suspects clothing?", or 'short' answer questions that requested a specific answer, such as "What colour was the suspect's shirt?". They found that the majority of questions used in the forensic interviews were of the short answer variety. These questions may have the advantage of eliciting information that the interviewer feels is forensically relevant and prevent the eyewitness from wandering off the point but they can also cause problems.

Fisher et al. found that short answer questioning appeared to produce a less concentrated form of retrieval. Eyewitnesses took less time to respond to short answer questions than for open-ended questions, which may be due (at least in part) to less time being spent actively trying to retrieve information. It was noted that both short-answer and open-ended questions were asked quickly of witnesses, thus there was a short latency between a question answer and the next question, giving no opportunity or encouragement to the eyewitness to elaborate or extend an answer. This use of short answer questions also changes the nature of the task from that of free recall. When short-answer questions are used the interview takes on the format of the interviewer asking short-answer question and the eyewitness giving a brief answer, the interviewer asking another short-answer question, and so on. This means that the interview changes from being directed by the eyewitness to being directed by the interviewer. Fisher et al. comment "It is difficult enough for the eyewitness to retrieve detailed events from memory when actively trying; it is virtually impossible when he remains passive" (p.181).

Using a question-answer format means that all the information elicited is that which is requested. Thus, if the interviewer forgets to ask a certain question, no information in that area is recorded, while if there were unusual occurrences during the crime, of which the interviewer is unaware, questions are not asked and information maybe omitted. As a guide, Fisher et al. suggest that most information should be gathered through the eyewitness's own free recall which should then be followed up with more specific questions later.

3. Inappropriate Sequencing of Questions.

The problems caused by the inappropriate sequencing of questions are similar to those associated with excessive use of question-answer format; both impair recall performance through shifts in attention. Fisher, Geiselman and Raymond (1987) noted that many of the questions asked by the interviewers were in a seemingly arbitrary order. They argue that this may impair eyewitness performance through shifting their retrieval efforts from one area to another. For example, if the interviewer asks a visually orientated question about the suspect's face, then follows with an auditory question about the suspect's voice, then returns to a visual target, such as the suspect's clothes, this shift in retrieval attention from one area to another and from one sensory modality to another may impair performance. Indeed, alternating retrieval across modalities has been shown in one study to produce a 19% decrease in eyewitnesses performance (Fisher & Price-Rouch, 1986).

Such decrements in performance are not limited to changes in sensory modalities. Shifts within a modality can also cause problems; for example, if one asks a visual question about the suspect's eyes, and then asks about the colour of the ceiling, the eyewitness must shift attention from one visual image to another. A more appropriate technique might be to gather all facial information at one time.

Further problems can be caused by asking what Fisher et al. term 'general knowledge' questions, such as "why do you think he did that?" or " Was he married?", in amongst questions concerning the crime. Shifting from the recall of crime details to general knowledge questions, then back to crime details can cause decreases in the eyewitness's performance.

4. Other problems.

Fisher, Geiselman and Raymond (1987) also identified some other problems that did not occur in all of the interviews that they recorded. They considered these problems to be less severe. The problems included negative phrasing, leading questions, inappropriate language, judgemental comments, lack of following potential leads, and an underemphasising auditory cues.

Negative phrasing occurs when questions are asked in the negative form. For example, "you *don't* remember if..?" Phrasing questions in this form may actively discourage the eyewitness from attempting to retrieve information in a concentrated manner. They occurred in many of the recorded interviews. Fisher et al. (1987) describe leading questions as, questions that subtly suggest that a certain answer is required. Not only are the demand characteristics of the situation likely to produce compliance, but Loftus (1979) has found that leading questions may actually bias eyewitnesses later recollections of an event.

Inappropriate language was found where interviewers used overly formal sentences or words, which were beyond the comprehension of the eyewitness. Such language may not only prevent the eyewitness from understanding the question, but also creates a barrier between the interviewer and the eyewitness which is not conducive to optimal performance. Judgemental comments were occasionally made often about the eyewitness's role in an incident. These may make the eyewitness defensive or may serve to offend the eyewitness, and it is difficult to see how they could enhance recall.

Fisher et al. found that Police officers in their study often failed to follow up on leads that they were given. They cite the case of one eyewitness who described a suspect as looking like a 'newspaperman'. There was no attempt to follow up such comments, to elicit why the eyewitness felt that the suspect looked like a 'newspaperman', that might produce a more objective description. They often found that auditory clues were underemphasised. The Police rarely enquired about what a suspect may have said or if they had an accent.

George (1991) in a field analysis looked at the performance of 28 British Police officers using their usual interviewing procedures. His findings may be considered to broadly support those of Fisher et al. (1987). This suggests that there is some consistency in police interview procedures which is surprising given that few officers receive training in interviewing (George, 1990). Thus the issues raised by Fisher et al. would appear to be widely applicable.

3.3 Conclusions concerning the standard police interview

The way in which eyewitnesses are interviewed is dependant on what information is available to the police. If a great deal of information is available such as confessions, fingerprints, etc, then little effort may be placed on interviewing eyewitnesses. However, if little information is available from other sources then considerable emphasis is likely to be placed on eyewitnesses' testimony. Because of the nature of police interviewing this means that the greater the emphasis that is placed on the testimony of an eyewitness, the less likely this information is able to be checked.

However, analysis of police interview procedures indicates that there may be considerable scope for improvements. But what, ideally, might a technique designed by psychologists be able to provide a forensic interviewer?

There seem to be two major considerations.

1. The technique should reliably enhance accurate eyewitnesses recall. Accurate recall should be increased without corresponding increases in inaccurate information.

2. The technique should produce testimonies in which eyewitness confidence is related to accuracy, such that the more confident an eyewitness is with respect to an item of information, the more likely that item is to be correct. This might aid the forensic interviewer considerably (c.f. chapter two).

With this as a background, we can now turn to methods of facilitating eyewitness memory.

CHAPTER 4

THE COGNITIVE INTERVIEW

4.1 Introduction to the cognitive interview

Recently, considerable interest has been shown in a procedure developed by Geiselman and Fisher to enhance eyewitness performance which they term the 'cognitive interview' (Fisher, Geiselman & Amador, 1989; Fisher, Geiselman & Raymond, 1987; Fisher, Geiselman, Raymond, Jurkevich & Warhaftig, 1987; Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian & Prosk, 1984; Geiselman, Fisher, MacKinnon & Holland, 1985; Geiselman, Fisher, MacKinnon & Holland, 1986; Geiselman & Padilla, 1988). Because all new recruits within the UK will eventually be trained in this interview procedure, it will be considered in detail here. The main aim of the present chapter is to evaluate the efficacy of this procedure in the light of the criteria outlined in chapter three, that is:

1. Does the technique reliably enhance accurate eyewitnesses recall? Correct recall should be increased without corresponding increases in incorrect information.

2. Does the technique produce testimonies in which eyewitness confidence is related to accuracy, such that the more confident an eyewitness is with respect to an item of information, the more likely that item is to be correct?

Geiselman and Fisher, together with various co-workers, have attempted a systematic approach to their work on enhancing eyewitness recall that has involved four main factors.

1. A comprehensive review of the 'cognitive' literature.

Geiselman and Fisher have evaluated the current theoretical and experimental literature with regards to memory in order to determine what could be usefully applied in the forensic situation.

2. A systematic analysis of real life forensic interviews.

They have undertaken an analysis of real-life forensic interviews to look at the way real interviews are conducted and how they might be improved. This is important as there appear to be some fairly obvious problems with standard forensic interviews that may be dealt with without recourse to complex psychological theories, these were outlined in the previous chapter.

3. A comprehensive experimental programme.

They have implemented a comprehensive experimental programme in an attempt to evaluate the relevant issues in a systematic and realistic way.

4. Reporting of findings.

They have reported their findings not only to psychologists, but also to those to whom the results are of most concern, Police officers. This has been achieved by publishing the results in Police journals, for example the 'Journal of Police Science and Administration', using language which is easily understood by non psychologists (e.g. Fisher & Geiselman, 1992). Indeed they report that the Police have been continually involved at all stages and the relevant data have been reported to them.

4.2 Theoretical and experimental Background: Context effects

Part of the development of the cognitive interview has involved taking the results from 'pure' cognitive psychology and applying them to the field of eyewitness testimony. Although there are a number of cognitive findings that are used in the cognitive interview (see Bekerian & Dennett, 1993; Memon & Bull, 1991), the area of 'context' effects has been particularly significant.

Context effects have been evident in the cognitive literature for many years. The basic assumption underlying these effects is that material will be remembered better if the recall or recognition of the material takes place in the same or similar context to that in which the material was learned. The idea is that reinstating elements of the learning context will provide the subject with retrieval cues to enhance memory. One example is an experiment by Godden and Baddeley (1975), which tested the memory capacity of deep-sea divers. Subjects were asked to learn lists of words, either on a beach or under 15 feet of water. Their recall was tested either in the same environment or in the opposite environment. Godden and Baddeley found that environment had no major effect if recall was conducted in the same environment as encoding. However, if the subjects encoded the information in one environment then were tested in the other, recall was dramatically impaired and subjects remembered approximately 40% less information. The results of Godden & Baddeley have been replicated by other researchers, though with less dramatic effects (e.g. Smith, Glenberg & Bjork, 1978). It could be that the effects are less dramatic because the differences between encoding and retrieval conditions are also less dramatic.

Smith (1979) investigated whether it is necessary to physically reinstate the same environment for context-dependency to work or whether it is sufficient simply to imagine the original environment. He had subjects study words in a distinctive basement room one day, then had them recall the words either in the same room or in a different fifth floor room the next day. Subjects in the basement recalled about 18 words, significantly more than the group tested in the fifth floor room who could only recall about twelve words. A third group was also tested in the room on the fifth floor; however, these subjects were instructed to remember as much as possible of the original learning environment of the room in which they learnt the words before they were required to try and recall the list. This group recalled an average of 17.2 words, which was not significantly different from the average score of those who were tested in the same physical environment, but significantly more than those who were simply tested in the fifth floor room. Hence it appears the original context does not necessarily have to be physically reinstated; simply thinking of the encoding context can enhance recall.

There is also evidence of context effects that are dependent on internal mood states. Goodwin, Powell, Bremer, Hoine and Stern (1969) tested the effects of alcohol on memory tasks. They found similar results to those of Godden and Baddeley; information encoded when subjects had drunk alcohol was more accurately recalled when subjects drank alcohol again than when they were sober, and vice versa. Similar findings have been found with happy and sad moods. Teasdale and Fogarty (1979) found that when mood was manipulated subjects who were sad found it easier to recall prior sad experiences than happy experiences.

Thus, there seems to be considerable support for context effects, including mental reinstatement, in the experimental literature. Consequently, this idea has been fundamental to the development of the cognitive interview.

4.3 The cognitive interview

The 'original' cognitive interview was an attempt to combine existing psychological knowledge with interviewing (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian & Prosk, 1984). The original cognitive interview procedure involved four main instructions to the subject or eyewitness.

1. Subjects were asked to *reinstate the context*. They were asked to try to reinstate, in their minds, the context surrounding the target incident. This involved thinking about what the surrounding environment looked like at the scene; for example, rooms, the weather, any nearby people or objects. They were also asked to think about how they were feeling at the time, and their reactions to the target incident.

2. Subjects were asked to *report everything*. They were informed that some people hold back information because they feel that it is not important. However, they were required not to edit their accounts, even if they felt that the information that they remembered was not important.

3. Subjects were asked to *recall the events in different orders*. They were told that it is natural to go over events from beginning to end, but were also asked to try to go through the event in reverse order, or to start with the thing that impressed them the most in the incident, then go from there, working both forwards and backwards.

4. Subjects were asked to *change perspectives*. They were asked to try to recall the incident from the perspective of other people that were involved in the incident. For example, they were asked to try to place themselves in the role of a prominent character in the incident, and think about what they would have seen.

Later, Fisher and his colleagues produced what they considered to be an improved version of the cognitive interview (Fisher, Geiselman, Raymond, Jurkevich & Warhafig, 1987; Fisher, Geiselman & Amador, 1989). This refined or 'enhanced' cognitive interview sought to redress certain problems that had been encountered with the original procedure, and to incorporate the findings of the Fisher, Geiselman and Raymond's (1987) study of Police interview techniques.

An important consideration addressed with the 'enhanced' cognitive interview is the structure of the interview. Although the original cognitive interview provided instructions at the beginning of the interview little advice was given about conducting the remainder of the interview. Specifically, no guidelines were given about the sequential structuring of the interview. As inappropriate structure and questioning may hinder efficient recall, an important aim of Fisher et al. when modifying the cognitive interview was to develop guidelines for the order of forensic interviews.

Many of the refinements to the cognitive interview will appear obvious in the light of Fisher et al.'s (1987) article described earlier. Essentially the 'enhanced' cognitive interview is the standard cognitive interview with additional instructions to ensure that the following are included:

1) Time is spent building rapport with the eyewitness. This is achieved by getting to know the eyewitnesses, trying to put them at ease, and ensuring that they are relaxed and aware that they will not be interrupted.

2) The interviewer structures the interview so that it is directed by the eyewitness, thus, allowing the eyewitness time to concentrate, and structuring the interview so that it is 'compatible with the mental operations of the witness' (Memon & Bull, 1991, p.295). The interviewer avoids fixed styles of questioning, tries to empathise with the eyewitness' mental operations and avoids interrupting the eyewitness by holding back questions where appropriate.

3) The interviewer helps the eyewitness to produce 'focused retrieval'. The interviewer must 'encourage and assist the witness to generate focused concentration' (Fisher, Geiselman & Amador, 1989, p.723), principally through motivation. Further, Fisher et al. (1989) recognised that retrieval is a difficult task

that requires motivation; thus they explicitly state that 'the effective interviewer must encourage the witness to make the extra effort' (p.723).

Thus, although the 'enhanced' cognitive interview is similar to the original cognitive interview, it differs mainly in that the eyewitness directs the content and direction of the interview rather than the interviewer. In this respect it also differs from conventional interviews.

4.4 Empirical support for the cognitive interview

Some initial support for the cognitive interview came from a study by Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian and Prosk (1984). This was an experimental investigation of the effectiveness of the original cognitive interview compared to control conditions. The study used 16 undergraduate psychology students, who were told that they would be taking part in an experiment to improve their memory. However, while apparently taking part in a memory experiment they witnessed a staged incident; this was the interruption of the 'experiment' by an argument. Subjects were tested for their recall of the argument, using a test booklet. The initial question was '..what do you remember of the incident involving the person (or people) who interrupted the experiment at our last meeting?' (p.76). Subjects were asked to put one piece of information on each line of the test booklet. Subsequently more 'pointed' (i.e. specific) questions were asked.

The experiment revealed that significantly more correct information was recalled with both the open, and the 'pointed' questions, with the cognitive interview. Moreover, this increase in correct information was achieved without a corresponding increase in incorrect information.

Aschermann, Mantwill and Kohnken (1991) conducted a partial replication of the Geiselman et al. (1984) study. This study used 29 German undergraduate psychology students. Subjects were shown a short film which was presented incidentally. They were tested between two and nine days later. Subjects were tested in a similar manner to the Geiselman et al. (1984) study. The results showed a significant overall increase in correct information recalled in the cognitive interview condition, compared to a standard interview condition, and this was especially apparent with the open-ended initial question. There was no significant difference between the cognitive interview and standard interview in the amount of incorrect information produced although there was a trend for more incorrect information to be produced with the cognitive interview, especially with open-ended questions.

It may be of some importance that both of the above experiments differed from a standard Police interview, in that rather than the interviewer listening to the eyewitness, then writing a statement, the subjects effectively wrote their own statements. This may have been advantageous in that the subjects may have felt less hurried and more able to think about and report details.

Geiselman, Fisher, MacKinnon and Holland (1985) conducted a follow-up experiment to their original study, using a larger sample of 89 undergraduate psychology students. Subjects were presented with one of four films that were used by the Los Angeles Police Department to train Police officers. In each film at least one individual is shot and killed. Subjects were made explicitly aware that they would be later be tested on the film. After approximately forty-eight hours subjects' recall for the film was tested. In this study interviews were conducted by law-enforcement personnel trained in the use of the cognitive interview, 'hypnosis' or using their own usual interviewing methods. By comparing the cognitive interview with the law-enforcement personnel's usual methods this study avoided the comparison of the cognitive interview with an artificial 'control' procedure devoid of many of the problems identified by Fisher, Geiselman and Raymond (1987).

By transcribing the information produced by the subjects, it was found that there was a significant, 40% improvement in the amount of information produced using the cognitive interview when compared with the standard interview, and a 30% increase with the use of hypnosis compared to the standard interview. There was no significant difference between the amount of incorrect information produced by the groups. This is noteworthy given that there is considerable concern that increases in incorrect information may occur with 'hypnotically' elicited information (c.f. chapter four).

In addition, George (1991) conducted an experiment using the 'enhanced' cognitive interview. Subjects in a lecture witnessed a staged incident. Two weeks later subjects were interviewed by Police officers using a) the 'enhanced' cognitive interview, b) a procedure of Conversation Management (a technique designed 'to equip interviewers in the social and communication skills required to open, and keep open, channels of communication in order to find out facts' p.3.), or c) standard Police interviews. The results of these interviews were transcribed-interviewers were explicitly told 'there is no need to capture the information in writing' (p. 97). There was a trend for the 'enhanced' cognitive interview to produce more information, but this was not statistically significant (this could have been due to the small sample size; only four subjects in the cognitive interview group and 15 in the whole experiment). There was no indication of increased errors or confabulations in the cognitive interview group.

Other experimental investigations have shown a similar, though variable, pattern of increases in recall for the original cognitive interview with no significant increases in errors (for a review see, Bekerian & Dennett, 1993).

However, Bekerian, Dennett, Reeder, Sloper, Saunders and Evans (1994) report that significantly more incorrect information was produced using a cognitive interview procedure than a standard interview. Thirty-seven undergraduates were played a film which they were asked to recall twenty-four hours later. A request for subjects to remember as much as they could, and to recall information in any order, was given to both the cognitive interview condition and the standard condition. In addition, subjects in the cognitive interview group were asked to use context reinstatement and to 'recall everything'. The cognitive interview group produced more correct information than the standard group, but they also produced significantly more incorrect information. Bekerian et al. (1994) state 'There are no immediate reasons as to why incorrect recall was effected by [the] cognitive interview' (p.3).

In addition to studies comparing the cognitive interview with various controls, Fisher, Geiselman, Raymond, Jurkevich and Warhaftig (1987) also conducted an experimental investigation comparing the cognitive interview with the 'enhanced' cognitive interview. A similar protocol to that of Geiselman et al.'s original 1984 study was used. Subjects were shown a video recording of a simulated violent crime then their memory for the video was tested 48 hours later.

Subjects recall was transcribed from tape recordings of the interviews. The study showed the 'enhanced' cognitive interview to produce 45% more correct information than the `origional' cognitive interview, an increase which was not accompanied by a corresponding increase in errors. Fisher et al. further coded the data produced by the 'enhanced' cognitive interview and the `origional' cognitive interview, to determine if the extra information produced by the 'enhanced' cognitive interview was simply trivial information. They found the increased information produced by the 'enhanced' cognitive interview' to have a similar proportion of information relevant to the crime as the `origional' cognitive interview. However, although this extra information recalled in the 'enhanced' cognitive interview was deemed **relevant** to the crime by the experimenters, on the basis of the coding scheme as reported, it is not clear whether information that the experimenters called relevant was similar to what an investigating officer would term relevant.

More recently a study has questioned the usefulness of the cognitive interview (Memon, Bull & Smith, 1995; Memon, Holley, Milne, Kohnken & Bull, 1994). Subjects witnessed a staged armed robbery and were subsequently interviewed by officers trained in what Memon et al. labelled a 'cognitive' or a 'structured interview'.

Initially, officers in both the cognitive interview and the structured interview were asked to begin the interview with the statement and free recall 'tell me in your own words what happened this morning'. There were no differences between the cognitive interview and the structured interview conditions in this control situation. Next, subjects were interviewed using cognitive interview or structured procedures.

No differences were found in terms of correct information, incorrect information or suppositions.

However, there are a number of possible explanations for these findings. Firstly, Fisher & Geiselman have stated that context reinstatement is most effective if context is reinstated before free recall; however, this was not the case in this experiment. It is plausible that reinstatement of context may be most effective given in conjunction with report all instructions at the beginning of an interview. Reinstatement of context later may not be as effective as subjects may have already established an interview routine and may rapidly tire of repeated retrieval attempts in a short period of time. Secondly, the interview that was termed a structured interview may not have been comparable with a standard police interview. While Fisher, Geiselman and Raymond (1987) and George (1990) have characterised standard police interviews as having many interruptions Memon et al. used a procedure that allowed initial uninterrupted recall. As Fisher et al. note, if a witness is interrupted (or not) in the initial stages of the interview, this sets the context for the way in which the rest of the interview will be conducted. Further, it could be argued that interviewers in the 'standard interview' condition were not given standard (non cognitive interview) police training. Instead, arguably, they were given instructions that overlapped considerably with the 'enhanced' cognitive interview but with out the four mnemonic instructions (i.e. establishing rapport, use of open questions, not interrupting the witness, and role-playing). This may be particularly important given that recently, Fisher has commented:

'My impression is that [the cognitive interview] works primarily because it facilitates communication and only secondarily because it improves memory retrieval' (McCarthy, 1993, p.28).

Nevertheless, the findings of Memon et al., raise some doubts as to the reliability of the four mnemonic strategies in enhancing recall.

This raises the issue of exactly what it is about the cognitive interview that is responsible for the increases in recall. Boon and Noon (1994) investigated the effectiveness of the report all, reinstate context, change order, and change perspectives mnemonics on subjects recall of a stimulus video film. Subjects were allocated to one of five groups that were required to make two successive recalls of the film. Three of the groups were requested to recall the film using first the report all mnemonic followed by a second attempt with either a change order, change perspectives or reinstate context mnemonics. The fourth group was asked to use the report all mnemonic for their first recall attempt but not for a subsequent attempt. The fifth group was not required to use any retrieval mnemonics (control).

All the mnemonic groups showed improved overall correct recall compared to the control group. Further analysis showed significant improvements in recall for test two compared to test one when reinstate context, change order, and try again (after being instructed to report all for test one) were used. However, no improvements were found when subjects were asked to change perspectives or try again without being given instructions to report all at test one. Thus, it would appear that the mnemonics reinstate context, report all, and different orders can enhance recall, while change perspectives does not appear to be as effective, at least in this situation.

4.5 The cognitive interview with children

Experimental investigations of the cognitive interview have also included various evaluations of its efficacy with children. The use of child witnesses has become a controversial issue recently because although children may potentially be able to provide useful information, their suggestibility and accuracy have been questioned. Clearly, therefore a technique to improve their accurate memory performance would be beneficial. Results have been mixed.

For example, Geiselman and Padilla (1988) used the original cognitive interview to test the memories of children, between seven and twelve years of age, for a video they were shown of a liquor store hold up. Their ability to remember details was tested three days later using either the cognitive interview or a standard interview. Children interviewed using the cognitive interview produced 21% more information than those tested with the standard interview without a corresponding increase in inaccurate information.

Similarly, Saywitz, Geiselman and Bornstein (1992) conducted a similar experiment but using a staged incident rather than a video film. They found similar effects; a 20% increase in information recalled for eight and nine year olds using the 'enhanced' cognitive interview compared a control group of similar age. A 44% improvement was found for 11 and 12 year olds, when compared to the appropriate control group. This performance was further improved to increases 25% and 66% respectively, over controls in the corresponding age groups, if the cognitive interview was practised before the test session. No increase in the amount of incorrect or confabulated information was observed.

In a further study, Dietze and Thomson (1993) compared the recall performance of six year olds, 11 year olds and adults, both with and without an abbreviated form of the cognitive interview. Their results showed an increase in the amount of information recalled with the cognitive interview when compared with the free recall condition for each age group. No significant corresponding increases in errors were found. The amount of information recalled also increased with age. Interestingly, Dietze and Thomson (1993) suggest that children's failure to perform as well as adults, is due to them encoding less information, so even with an optimal retrieval strategy they would not perform as well as adults. Nevertheless they suggest:

> 'If the performance differences between children and adults only reflect children's problems in utilizing an appropriate retrieval plan, then one would expect that children would benefit more from the use of an appropriate retrieval plan than would adults' (p.105).

However, others have found problems using the cognitive interview with children (see Milne & Bull, 1994). For instance, Milne, Bull, Kohnken and Memon (1994) conducted a study with eight and nine year olds testing their recall for a video which they had seen. They found a significant increase in both correct and incorrect information recalled in the cognitive interview group. Kohnken, Finger, Nitsche, Hoefer and Ascherman (1992) found a similar pattern of results. Correct information was increased by 93% but this was also accompanied by a significant increase in incorrect information. On the other hand, Memon, Cronin, Eaves and Bull (1992) found no significant differences between a standard interview and a cognitive interview. Furthermore, an analysis of the interview transcripts from Memon et al.'s study has indicated that not all children understood or used the Cognitive interview technique (Cronin, Eaves, Kupper, Memon & Bull, 1992).

4.6 Problems of ecological validity

On the basis of these experimental studies the utility of the cognitive interview appears to have the potential to increase accurate recall in the majority of cases. However, there is some suggestion that it may be accompanied by increases in confabulation in some situations, perhaps especially with children.

An obvious criticism, however, of many the cognitive interview studies is that they lack ecological validity; i.e. they are too artificial. Geiselman et al. (1985) make claims about their stimulus materials such as, 'The scenarios are realistic in that monitored reactions of officers in training have been found to be comparable to reactions that would be expected in similar street situations' (p.404). A thorough evaluation of this assertion would require further information about the empirical work that forms the basis of these conclusions. It is not obvious, for example, that a film clip showing a killing will produce similar reactions to a real-life killing. Also, unlike in some of experimental studies, in 'real' life situations, of course, eyewitnesses are not usually aware that they may be tested later, so they would be unlikely to encode the information regarding the target event as effectively as in a situation where testing is anticipated. Even when subjects are not explicitly told that they will have to recall stimuli material later, the experimental situation is usually such that they deduce that they will be required to recall, or will be asked questions about the stimuli later.

Fisher et al. (1989) have noted that the results of their experiments would always be questioned until they are demonstrated in the real world. They state '...if the cognitive interview is to be applied outside the friendly confines of the laboratory, it must be demonstrated to be effective in the real world' (p. 724). Fisher et al. (1989) therefore set out to investigate the 'enhanced' cognitive interview in a field setting. George (1991) shortly afterwards, conducted a similar field investigation.

4.7 The cognitive interview in the field

On first consideration, the Fisher et al. (1989) and George (1991) studies appear to provide some of the most convincing evidence to date of the effectiveness of the cognitive interview. The method used in the field study by Fisher et al. (1989) was to tape record interviews of eyewitnesses to real crimes. The interviews were conducted by 16 experienced detectives from the robbery division of a Police force in Florida, USA. Preliminary recordings were made of interviews conducted by the detectives on eyewitnesses, before any training in the cognitive interview was given. In all, 88 interviews were recorded before training, and 47 interviews were conducted after training; 24 in a group using the cognitive interview and 23 in a control group.

The interviews that were used were selected according to the following strict criteria. 1) The case had to be severe enough that time would be made available to conduct a thorough interview. 2) The eyewitness must have had a 'decent' chance to observe the incident and the suspect. 3) The eyewitness had to be fluent in English and also be co-operative.

Fisher et al. (1989) state that some interviews were rejected as unsuitable because of reasons such as, the eyewitness was intoxicated, the interview was a couple of days after the incident, the suspect was known to the eyewitness, or a suspect had been detained for identification. The last reason may appear strange but when a suspect is in custody, Police tend to take a less comprehensive attitude to interviews, preferring to secure a positive identification instead. (Although rejection for these reasons would clearly create a more homogenous sample of interviews; in the case of intoxication or delayed recall one might have expected the cognitive interview, with its emphasis on reinstating a different context, to perform even better than in the other conditions.) Seven detectives completed the training programme. The recording of post-training interviews took seven months to complete and the tape recordings were then transcribed by research assistants.

Fisher et al. (1989) evaluated the effectiveness of the cognitive interview in two ways; 1) by comparing the number of 'facts' elicited before and after training in the use of the cognitive interview, and 2) by comparing the number of facts elicited by the trained detectives using the cognitive interview and the control group of detectives who were still using standard techniques. When detectives who were not going to be trained in the cognitive interview were compared to the detectives who were to be trained in the cognitive interview there were no significant differences between the two groups. However, after training in the cognitive interview there was a significant improvement; 63% more information was recalled by eyewitnesses interviewed by detectives trained in use of the cognitive interview compared to those interviewed by the 'control' detectives. Moreover, detectives in the cognitive interview trained group showed a 47% increase in the amount of information that they elicited from eyewitnesses compared to their previous performance before training.

However, whilst these results seem impressive, there are a number of difficulties in the interpretation of this study. One possible problem concerns the way in which statements were scored. Fisher et al. did not score opinionated responses, such as comments like "the guy seemed nervous" (p.724). It is possible that this scoring method may have lead to artificially high performance in the cognitive interview group. Because of the explicit instructions to try harder (and perhaps implicit instructions to be more confident), comments which would not have been scored such as "the guy seemed nervous" in the Police interview may have changed to a comment like "the guy was nervous" with the cognitive interview and have been scored accordingly.

It is also notable that of the seven detectives trained in the cognitive interview, one (10%) produced a **decrease** in performance of 23%. Clearly, on the basis of such a small sample it is difficult estimate whether this was a curious anomaly or whether this represents a potential problem. Fisher et al. (1989) comment of this detective: 'Not coincidentally an analysis of the post-training interviews showed that he was the only one of the seven detectives who did not incorporate the recommended procedures into his post-training interviews' (p.724). However, despite this explanation, this example does raise further questions. How and why did this Police officer fail to incorporate the recommended procedures into his interviews? Presumably, he passed the training procedure, including a practise interview in the field, and received individual feedback on his performance, as outlined in Fisher et al.'s (1989) method section. Did his performance improve or become worse over the repeated attempts? As such a result has not been reported in previous studies in the laboratory or experimental situations why did it occur in the field? This example suggests that if the cognitive interview is to be used widely, the performance of individual interviewers should be carefully monitored. Indeed, George (1991) has also noted that some officers pick up the cognitive interview better than others, and suggests that training should be concentrated on certain individuals.

Another problem of interpretation concerns how Fisher et al. (1989) tackled the question of accuracy. Obviously, in a field situation accuracy is difficult to determine as there is often no way of definitively establishing what actually occurred. Fisher et al. (1989) therefore estimated accuracy by comparing each eyewitness report with that of what they term another 'reliable' source, when this was possible. In 22 cases this source was another eyewitness, in one case a confession, and in one case information was supplied by a video camera. Fisher et al. (1989) found there to be a 93% corroboration rate with information produced by a 'reliable' eyewitness for information produced by detectives untrained in the cognitive interview, and a 94.5% corroboration rate for detectives using the cognitive interview. Fisher et al. (1989) note that their corroboration levels ares high when compared with the accuracy levels typically produced by laboratory studies, and they cite the similar findings of Yuille & Kim (1987). They state:

> "If this difference between laboratory and field studies continues to appear, one may question the validity of describing in court the accuracy rates found in the laboratory as evidence of the general unreliability of eyewitness testimony in field cases" (p.725).

However, corroboration rates in field studies and estimates of memory accuracy in experimental studies may be quite different measures. There could be a marked difference between the accuracy of corroborated information and the accuracy of all information produced. By definition corroborated information is information which two or more eyewitness have recalled. Therefore one can assume such information is probably *central* to the eyewitness situation and more likely to be recalled by several eyewitnesses. Laboratory situations typically consider both central and perhaps less important *peripheral* information to determine accuracy rate. If only the *central* information from laboratory studies were considered, the accuracy rate would probably be closer to that of the corroboration of the Fisher et al. (1989) and Yuille and Kim, (1987). Also as Fisher et al. (1989) themselves note, just because two items are correlated by two eyewitnesses, does not necessarily mean that they are accurate -both may be wrong.

However, perhaps the most important problem concerns how we identify the elements in the cognitive interview responsible for the reported improvements. For example, as already pointed out, the group trained in the cognitive interview produced a significant increase of 47% more information using the cognitive interview, compared to their previous performance. Of the seven detectives, the range of improvement, for six of them, was between 34-115%. What exactly was responsible for this range? How did the detective who produced a 115% increase in the number of facts elicited manage this? Did he reinstate context better than the others? Was his pre-training performance especially bad due to excessive interruptions? Did the training motivate him to try harder to motivate his own eyewitness? Without answers to such questions it is difficult to determine whether the improvements were a consequence of the cognitive interview per se, or to more generally features of the situation such as 'training' per se, or the fact that the situation was novel or different. The detectives were aware that they were evaluating a new technique, which they had been specially trained to use. The training for, and use of, a new technique may in itself have produced positive improvements in the Police officers' performance, by motivating them to try harder and in turn, to motivate the eyewitnesses to try too. Whilst it could be argued that this effect may be desirable no matter how it comes about, it may fade as the 'novelty' of the technique wears off, and may have implications for what training should entail.

Of relevance here therefore is the field investigation in the United Kingdom by George (1991). In this study, 28 Police officers were evaluated in one of four conditions; seven in each condition. A recording of an interview performed by each officer was evaluated before each was trained in an interview technique or placed in the control group. The interview techniques were: 1. the 'enhanced' cognitive interview; 2. Conversation Management; 3. Conversation Management combined with the 'enhanced' cognitive interview and 4., a control group. The results indicated that the 'enhanced' cognitive interview showed an improvement when compared to the standard Police interview control group of 14% more information. When compared to performance before 'enhanced' cognitive interview training this improvement was 55%. This advantage was for all kinds of information (i.e. who, what, when, where, how and why). Neither the Conversation Management nor the combination of Conversation Management and 'enhanced' cognitive interview produced more information than the untrained group. These results would suggest that it was not 'training' *per se*, or novelty alone that accounted for the improvements that occurred with the cognitive interview.

Nevertheless, a major finding in George's study was, that of the four mnemonic strategies suggested in the cognitive interview, three were hardly utilised. The instruction "not to edit anything out" was only minimally employed. The two other mnemonic aids which were rarely used were the instructions for a 'change of perspective' and a 'change of order'. George notes that it is unsurprising that officers rarely used the change of perspective mnemonic as, 'it is not an easy concept to ask someone to put themselves in someone else's shoes to review an event asking them to say what they think they would have seen, and remain confident that there will be no confabulation' (p.117). Critics have also suggested that the use of the 'change of perspective' mnemonic may make it difficult to use such statements in court, especially if children are interviewed, because of a danger of confabulation. The Police officers in this study may have had an intuitive grasp of this and so, this may explain why they did not use the technique. Some research into why these three techniques were not used and the implications for their inclusion in the cognitive interview would seem is appropriate. It is not clear if a similar pattern of mnemonic usage was present in the Fisher et al. (1989) study.

The fourth mnemonic, reinstatement of context, was widely used, apparently to great effect. This led George to conclude that 'where contextual reinstatement was present more information was elicited regardless of question type' (p.118). This effect was consistent across the seven types of information into which George coded recall. Reinstatement of context thus appears to be the most reliable technique for increasing recall. Why, however, the effects of the cognitive interview should be eliminated when it is combined with another procedure (Conversation Management) remains somewhat of a mystery.

Although it is time-consuming to perform, the field study clearly has been an effective paradigm in this area. And despite the problems in interpretation, it is very notable that the improvements found in field studies are remarkably similar to those which have been found in previous laboratory based experiments. Indeed, it could be argued that the cognitive interview might work *more* effectively in the field than in artificial experimental situations, because of greater contextual disparity between encoding and recall contexts in real life. In the laboratory, both encoding and retrieval contexts are usually fairly similar; for example, studies are often conducted in college rooms in a situation inducing low emotional arousal. In real eyewitness situations the differences between encoding and retrieval conditions are likely to be markedly different. For example, an individual may be drunk and frightened in a street at the encoding of an event but sober and relaxed in a Police interview room when they are required to remember the event. Assuming the mental context reinstatement is effective, one might therefore expect the cognitive interview to show even greater performance increments over standard interviews in these latter circumstances.

4.8 Confidence, accuracy and the cognitive interview

Despite this volume of research on the cognitive interview, very little work has been conducted to investigate effects of the cognitive interview on C-A relationships. This may be because of the general acceptance amongst experts in this area that there is little relationship between eyewitnesses' confidence and their accuracy (c.f. chapter two). Indeed, Fisher & Geiselman make this point in their book for investigative interviewers (Fisher & Geiselman, 1992) in a section entitled 'increasing eyewitness confidence'. In fact they go so far as to suggest that interviewers should try to increase the confidence of eyewitnesses, stating that interviewers shouldn't worry that this may increase confidence in false information as there is little relationship between confidence and accuracy anyway (p.38).

In one of the few studies that considered C-A relationships (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian & Prosk, 1984) subjects were asked to rate their confidence in each piece of information that they provided concerning a video film that they had seen. Subsequently more 'pointed' (i.e. specific) questions were asked. Confidence in correct information was significantly higher in the cognitive interview than in the control interview. Eyewitness confidence in incorrect information was also higher in the cognitive interview condition, but this was not significantly greater than that for the control conditions; it can be noted, however, that only a limited number of subjects were used, and there were few incorrect responses. It is unfortunate that when Aschermann, Mantwill and Kohnken (1991) replicated this experiment they did not include confidence judgements.

In another study Bekerian, Dennett, Reeder, Sloper, Saunders and Evans (1994) tested subjects memory for a crime film that they had seen. Subjects were interviewed 24 hours later with either a modified cognitive interview or a standard procedure. Forty-eight hours after seeing the video subjects were asked to give confidence ratings in information that they had provided in the previous interview. No differences in confidence between interview conditions were found. However, it is difficult to draw conclusions from this experiment as the cognitive interview instructions had been given 24 hours earlier so may not have been as effective as if they had been given immediately before subjects were asked to rate their confidence.

In sum, there has been very little work conducted that concerns C-A relationships with the cognitive interview, though it is plausible that the cognitive interview may enhance C-A relationships. Deffenbacher's optimality hypothesis suggests that the degree of 'optimality' of the information-processing conditions present at encoding, storing and retrieval influences C-A relationships (Deffenbacher, 1980). He suggests that under ideal circumstances witnesses will accurately gauge the adequacy of their memory performance in their memory reports while under conditions that are not optimal, the two will covary less reliably. Therefore the cognitive interview, by enhancing retrieval conditions, may help subjects to gauge their accuracy more effectively and so enhance C-A relationships.

4.9 Conclusions concerning the cognitive interview

Earlier two criteria were outlined by which memory facilitation interview procedures might be evaluated; 1) does the technique increase reliable recall? And 2), does the technique produce testimony in which confidence is related to accuracy? To what extent does the cognitive interview satisfy these criteria?

Although there are some possible problems with the methods of scoring responses which could have inflated cognitive interview performance, the vast majority of the studies described here have shown a substantial increase in recall with the cognitive interview compared to control conditions. However, a few significant increases in incorrect information have also been reported as well as some non-significant trends in this direction.

There is some evidence that confidence in correct answers has been significantly increased by the cognitive interview, whilst no significant increases in confidence in incorrect answers have, to our knowledge, been reported. But as so few studies have addressed the issue of the cognitive interview's effects on confidence/accuracy relationships, no definitive statement can be made at present with regard to this issue. In view of the potential importance of C-A relationships some more research on this issue would seem to be a priority.

CHAPTER 5

HYPNOSIS AS AN INTERVIEW TECHNIQUE FOR EYEWITNESSES

5.1 Introduction

Another method for enhancing accurate eyewitness recall that has received considerable attention is that of hypnosis. During the late seventies and early eighties this technique was associated with a number of successful, high profile cases. For example, in 1976 in Chowchilia, California, 26 children and the driver of their school bus were kidnapped at gun-point by three masked men. They were taken to a quarry and buried 6ft underground. Hours later, they were able to dig themselves out and were rescued. Despite questioning many of the victims, the FBI could not elicit specific descriptions of the suspects or other relevant details. Ed Ray, the driver of the bus, agreed to a hypnosis interview and was able to recall all but one digit of the number plate of the suspect's white van. This led to the capture of three males who were subsequently convicted of the crime (cited in Orne, Sokis, Dinges & Orne, 1984).

Such dramatic results in the field has led to suggestions that hypnosis can enhance eyewitnesses memory for events that they have witnessed by up to 60% (e.g., Haward & Ashworth, 1980; Hibbard & Worring, 1981; Kleinhauz, Horrowitz & Tobin, 1977; Reiser, 1976, 1980; see also Gudjonsson, 1992); moreover, Vingoe (in press) found that 40% of a sample of 'experts' on forensic hypnosis agreed that 'the accurate recall of useful material by an eyewitness to a crime is better during hypnosis than during his/her normal state of wakefulness' (see also, Kassin, Ellsworth & Smith, 1989). Clearly, if such claims were justified then 'hypnosis' would be a very valuable interview technique.

Indeed, the idea that hypnosis can enhance normal memory appears to be widely held belief amongst the lay public in many different cultures (Erdelyi, 1994; Wagstaff, 1993). This belief appears to extend to jurors' evaluation of hypnotically elicited testimony. Wagstaff, Vella & Perfect (1992) gave subjects identical trial transcripts. Subjects were divided into three groups; one group was told that the eyewitness statement was given using 'hypnosis' while the other two groups were told that the eyewitness statement was given using a guided memory procedure or free recall. When asked to decide whether the accused was guilty or not guilty, the subjects in the hypnosis condition were significantly more likely to find the defendant guilty than in the other conditions.

However, as a result of problems associated with the forensic hypnosis the Home Office guidelines of 1988 now discourage its use by police as a memory facilitation procedure in most cases (similar legal directives have been issued in the United States, see for example, Diamond, 1988; Orne, Sokis, Dinges, & Orne, 1984) Nevertheless, in a letter accompanying the Home Office guidelines the Association of Chief Police Officers has argued that hypnosis has been useful, and may be legitimately used in serious cases where few leads are available (see also, Gudjonsson, 1992). The evidence for and against the use of hypnosis as a memory aid can again evaluated according to the criteria that the technique should reliably enhance accurate eyewitnesses recall (i.e. accurate recall should be increased without corresponding increases in inaccurate information), and the technique should produce testimonies in which eyewitness confidence is related to accuracy, such that the more confident an eyewitness is with respect to an item of information, the more likely that item is to be correct.

To do this, it is first useful to give a brief introduction to the nature of hypnosis and hypnotic interviewing.

5.2 Hypnotic interviewing

Hypnotic interviews vary from situation to situation depending on the training and personality of the hypnotist and characteristics of the witness. However, in most cases a similar procedure is used (see, Reiser, 1980, 1990). The hypnotist establishes rapport with the witness and explains what will happen. Next a hypnotic induction procedure is used. Typically the witness is asked to concentrate on a particular physical or mental activity (e.g., staring at a target affixed to a wall) while suggestions are given (e.g. your eyelids are becoming heavy). After this initial stage the hypnotist attempts to 'deepen' the eyewitness's level of hypnosis. This can be done in a variety of ways, for example, counting may be used where it is suggested that the subject will go into a deeper and deeper state of relaxation with each count. Another method is to ask the eyewitness to imagine that they are in a very relaxing setting, for example, lying on a beach in the sun (Gudjonsson, 1992; Reiser, 1980).

After these induction procedures have been completed the hypnotist attempts to enhance recall. A variety of techniques may be used to achieve this. Two commonly used techniques are 'age-regression' and the 'television-technique'.

Age-regression involves asking the witness to go back in time and re-experience the original event. The television-technique involves asking the witness to imagine that they are watching an imaginary television screen on which they can watch the events that they had witnessed. The idea of this technique is that it may aid the recall of a traumatic event by distancing the witness from the event, while asking the eyewitness to relive the event may prove too traumatic for them. In addition memory facilitation mnemonics similar to those of the cognitive interview may be used (Wagstaff, 1982).

After the witness has finished recalling the event, further instructions may be given to the eyewitness to suggest that his/her memory will improve in time and over future sessions, prior to the witness being woken. Typically the hypnotist counts while suggesting that the subject becomes more and more awake.

5.3 The nature of hypnosis

While researchers agree what occurs in a forensic hypnosis interview, there is considerable controversy over what hypnosis is. The main controversy has been termed the state-non-state debate (e.g., Barber, 1969; Lynn & Rhue, 1991; Wagstaff, 1987, 1991, 1995), and this has a number of implications for research on hypnotic interviewing. Broadly, speaking most theorists can be construed as either state or non-state theorists. While most researchers of forensic hypnosis have tended to take a somewhat pragmatic approach to theories concerning hypnosis (i.e. they have focused on what influence hypnosis has on subjects' performance rather than theoretical explanations of this performance), these different theoretical preferences are frequently manifested in the way research is designed and interpreted.

State theorists argue that hypnosis produces profound alterations in the subject's state or condition (e.g. Barber, 1991; Bowers, 1983; Evans, 1991; Hilgard, 1986, 1991; Nash, 1991; Orne, Dinges & Orne, 1986; Sheehan & McConkey, 1982). One of the most influential theories of hypnosis amongst state theorists is that of Hilgard (1986). Hilgard argues that there exist in the mind multiple systems of control, or 'parts' of the mind, that are not all conscious at the same time. Normally these cognitive control systems are under the influence of a central control structure, or 'executive ego', that controls and monitors the other systems; but when the subject enters hypnosis the hypnotist takes over some of the normal monitoring such that, for example, in response to suggestion, motor movements (such as arm lowering) are experienced as involuntary (because the part responding to the hypnotist and actually controlling the movement is 'dissociated' from awareness), pain is reduced or eliminated (because the painful sensations are dissociated from awareness), and memory and perceptions are distorted such that, for instance, suggested hallucinations and false memories are perceived as real (because the part that would normally monitor and detect distortion is dissociated from awareness). In addition, according to the dissociationist approach, the hypnotist is to a certain extent able to access or talk to the different dissociated 'parts', and bring them into awareness. For example, although when given a noxious stimulus a hypnotic subject may appear to feel no pain the hypnotist may be able to speak to another 'part' of the subject's mind that does not feel pain (Hilgard, 1980).

In contrast, supporters of the non-state view reject the traditional notion of hypnosis as an altered state of consciousness and contend that various hypnotic phenomena are more readily explicable in terms of interactions between more mundane psychological processes such as imagination, relaxation, role-enactment, compliance, conformity, attention, attitudes and expectancies (see, for example, Barber, 1969; Coe & Sarbin, 1991; Sarbin & Coe, 1972; Spanos, 1982, 1986, 1991; Spanos & Chaves, 1989; Wagstaff, 1981, 1986, 1991, 1995). Wagstaff (1995) characterises hypnosis, from a non-state socio-cognitive or cognitive behavioural perspective, as a cultural invention or fantasy, that can be seen primarily as a strategic role enactment. In other words, the hypnotic subjects enacts the role of a 'hypnotised person' as defined by cultural expectations and cues provided by the immediate situation. The subject does not fall into a sleep-like altered state of consciousness, and does not lose control of his or her actions. Rather, hypnotic subjects adopt strategies to make suggestions come about; for example, in response to an hallucination suggestion subjects may try hard to imagine objects or events; to appear amnesic, subjects may employ distraction or inattention strategies; to reduce pain subjects may try to relax, or distract themselves. Non-state theorists claim that these strategies may be remarkably effective even in the reduction of surgical pain (see for example, Chaves, 1989), however, when the strategies are not effective, some non-state theorists contend that subjects may recourse to behavioural compliance or faking (Spanos, 1991, 1992; Wagstaff, 1981, 1986, 1991, 1995).

In effect then, to non-state theorists, terms such as 'hypnosis' or 'hypnotic' are simply labels that refer to situations defined by participants or observers as such (because, for example, the situations contain 'hypnotic induction' rituals); they do not refer to the existence of some kind of special state or process (Wagstaff, 1995). In recent times state theorists, faced with evidence that psychological processes such as imagination, relaxation, role-enactment, compliance, conformity, attention, attitudes and expectancies, influence hypnotic responding have acknowledged many of the arguments of non-state theorists, but still contend that a separate state of consciousness also is present (Bowers & Davidson, 1991; Hilgard, 1986).

The main implications of this debate for hypnotic interviewing would seem to be, only state theorists would predict that 'hypnosis' has some special capacity to improve memory compared to nonhypnotic memory facilitation procedures; non-state theorists would predict that if improvements in memory do occur with hypnosis they will result from nonhypnotic factors that may be present in the situation. However, importantly, both schools of thought might predict some degree of memory distortion with hypnosis; state theorists because the hypnotic state can evoke delusions and false memory, and nonstate theorists because the hypnotic situation contains social pressures that may encourage distortion (Wagstaff, 1993).

5.4 Empirical support for hypnosis as an eyewitness interview technique

While investigations of the cognitive interview in the laboratory and the field have shown broadly similar findings, the hypnotic literature tends to show differences between field and laboratory settings (Smith, 1983; Wagstaff, 1989). Impressions from the field suggest that hypnosis may be useful (e.g., Haward & Ashworth, 1980; Kleinhauz, Horrowitz & Tobin, 1977; Hibbard & Worring, 1981;

Reiser, 1976, 1980) while laboratory studies have been less supportive (Wagstaff, 1984; 1993).

Erdelyi (1994) categorised the experimental studies into four types according to the nature of the stimuli that were presented to subjects and the method of testing. Erdelyi categorised stimuli as being either 'low-sense' (e.g. inkblots, nonsense syllables, digits) or 'high-sense' (e.g. filmed crimes, staged incidents). Testing was categorised either as stimuli recognition or as stimuli recall (see also, Erdelyi, Dinges, Orne, Whitehouse & Orne, 1987).

Using these categories to distinguish between different studies, a broad pattern emerges. No hypermnesia is shown for recognition of low-sense stimuli (e.g., Young 1925), or recognition of high-sense stimuli (e.g, Putman, 1979; Sanders & Simmons, 1983; Sheehan & Tilden, 1983, 1984; Timm, 1981; Wagstaff, 1982; Wagstaff, Traverse & Miller, 1982; Zelig & Beidleman, 1981) or recall of low-sense stimuli (Barber & Calverly, 1966; Das, 1961; Dhanens & Lundy, 1975; Eysenck, 1941; Huse, 1930; Mitchell, 1932; Rosenthal, 1944; Salzberg & De Piano, 1980; White, Fox & Harris, 1940; Young, 1925).

However, when **recall** of high-sense material is considered hypnosis there are some reports that hypnosis may enhance recall performance (e.g., Cooper & London, 1973; Crawford & Allen, 1983; De Piano & Salzberg, 1981; Dhanens & Lundy, 1975; Dorcus, 1960; Gheorghi, 1972; Rosenthal, 1944; Sears, 1954; Sheehan & Tilden, 1984; Stager & Lundy, 1985; Stalnaker & Riddle, 1932; White, Fox & Harris, 1940). However, not all studies show significant effects (e.g. McConkey & Nogrady, 1984; Nogrady, McConkey & Perry, 1985; Timm, 1981; Wagstaff & Sykes, 1983; Wagstaff & Mercer, 1993), and none of the above studies controlled for response criteria. For while hypnosis may increase the amount of information that subjects provide in terms of correct information, it has also been associated with increases in incorrect information (e.g. Dwyan & Bowers, 1983; Erdelyi & Kleinbard, 1978; Klatzky & Erdelyi, 1985; Orne, 1979; Orne, Whitehouse, Dinges & Orne, 1988; Smith, 1983; Whitehouse, Dinges, Orne & Orne, 1988). Nevertheless, just as increases in hits do not necessarily mean that memory has increased (Klatzky & Erdelyi, 1985), increases in false alarms do not necessarily mean that memory has become distorted. Both may be explained by a lowered criteria for report consistent with Signal Detection Theory (Green & Swets, 1966).

Indeed, when one considers the social situation that is created when hypnosis is used it is not surprising that more information may be produced. Subjects are given a lengthy induction procedure and then memory is tested. In such situations, even if explicit instructions are not given (though they often are) it is obvious to the subject that hypnosis is expected to enhance recall. The subject may try to fulfil these demands by lowering their criteria for report and reporting more information.

To control for response criterion, Dinges, Whitehouse, Orne, Powell, Orne & Erdelyi (1992) and Whitehouse et al. (1988) compared hypnosis to a control procedures using a forced-choice response paradigms. These studies indicated that when productivity was controlled for, hypnosis did not enhance memory compared to the control group. Indeed Dinges et al. and Dywan and Bowers (1983) found that when the proportion of correct to incorrect responses were considered, subjects in the hypnotic condition performed worse than the control group.

5.5 Confidence and hypnosis

There is considerable evidence that there is little C-A relationship in eyewitness testimony when hypnosis is used (Orne, Soskis, Dinges & Orne, 1984; Sheehan, 1988). In his review, Sheehan notes that regardless of their accuracy hypnotic subjects express high levels of confidence in information that they provide. Furthermore, he notes that despite varying different methodologies and test situations the effects of hypnosis on confidence are remarkably consistent. Although not all the data support a weak or negative relationship, the evidence certainly suggests an absence of a large, positive relationship.

Research using Loftus-type paradigms (Loftus, 1979), based on the procedure of injecting misleading information subtly into a test situation well before testing, has been shown to increase hypnotic subjects' confidence in their memories including inaccurate responses, especially for subjects who were highly hypnotically susceptible (Sheehan, Grigg & McCann, 1984; Sheehan & Tilden, 1983, 1984, 1986). Dywan & Bowers (1983), using a similar paradigm to the work of Sheehan et al., also found that hypnosis increased subjects' confidence without any increases in accuracy. Furthermore, this effect was most apparent in subjects of high hypnotisability. Highly hypnotisable subjects, even if in parts of the experiment that were performed before hypnosis, displayed higher confidence in their answers than their low hypnotisable counterparts. These findings are also consistent with the work of Zelig and Beidleman (1981) and Rainer (1983). Again these findings are explicable in terms of the implicit or explicit pressure that the hypnotic process carries with it, i.e. that memory will be improved. As subjects appear to believe that a good witness is a confident witness, then subjects in hypnotic conditions would be expected to feel under pressure to increase their confidence (c.f. Brigham & Wolfskeil, 1983; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Stuve, 1988; Cutler, Penrod & Thomas, 1988; Fox & Walters, 1986; Leippe, Manion & Romanczyk, 1992; Lindsay, 1994; Lindsay, Wells & O'Connor, 1989; Lindsay Wells & Rumpel, 1981; Wells, Ferguson & Lindsay, 1981; Wells, Lindsay & Ferguson, 1979). Indeed this hypothesis is supported by the findings of Redston and Knox (1983) and Sheehan and Tilden (1984) who found that subjects that simulated hypnosis (i.e. were told to pretend to be 'good' hypnotic subjects) expressed even greater confidence in answers that they provided than the hypnotic subjects. Thus, the simulating subjects appear to have hypothesised that in order to appear like a 'good' subject they had to show high confidence.

However, not all the experimental literature supports the contention that confidence increases substantially with hypnosis. Redson and Knox (1983) found no increase in confidence for hypnotic subjects compared to controls. Sanders and Simmons (1983) also found that hypnotic subjects were not more likely to express greater confidence than controls on a measure of their 'willingness to testify in court'.

In sum, it would appear that hypnosis may, in some situations, increase confidence in both correct and incorrect information. However, in considering the influence of hypnosis on C-A relationships almost all studies have focused on the presentation of misleading post-event information and/or leading questions. Thus, as noted in the previous chapter, these findings may reflect the current negative view of eyewitness memory and may be more a reflection of experimental design than of eyewitnesses' abilities (Yuille & McEwan, 1985). It is plausible that if within-subject C-A relationships were calculated as suggested by Smith, Kassin and Ellsworth (1989) and biased instructions and misleading information are not used, then the C-A relationships may be higher than has been previously reported.

5.6 Conditions associated with forensic hypnosis.

Clearly, there are major discrepancies between findings reported in the experimental literature and those expressed by proponents of hypnosis with experience in the field. However, there would appear to be a number of factors present when hypnosis is used in the field that may contribute to an apparent enhancement of memory when it is used.

The experimental work that has been outlined illustrates that hypnosis may lead eyewitnesses lower their criteria for report with a corresponding increase in the proportion of inaccurate information. Furthermore, it has be suggested that confidence in information both correct and incorrect may be increased. As the amount of information that eyewitnesses provide, even if it is trivial, and their confidence in that information are used to judge how good eyewitnesses are, hypnosis may *appear* to investigating officers to have enhanced eyewitness performance (Bell & Loftus, 1988, 1989; Brigham & Wolfskeil, 1983) especially when little of what they say may be verified. Also, if eyewitnesses are more susceptible to leading questions with 'hypnosis' than standard procedures, one would expect the replies to leading questions to be more in line with what the police expected than with standard procedures. Because the police receive more information that supports their beliefs about what occurred than with a standard procedure they may accept that information uncritically, in a positive way and so have a more favourable impression of hypnosis.

Also, when 'hypnosis' is used in forensic situations there may be even greater pressure on witnesses to increase the amount of information that they produce and express false confidence than there is in the laboratory. The very fact that the police are using the technique implies that it is effective. This combined with instructions from the 'hypnotist' that the technique will enhance recall may encourage the eyewitness to believe that what they imagine or vaguely recollect are accurate memories. They may also lower their criterion of report. Hypnosis is usually used as a last resort, when other methods have failed and there are no other leads. The eyewitness is likely to be aware of these circumstances before agreeing to be 'hypnotised'. This, combined with the knowledge that the police have gone to a great deal of trouble to use a 'special technique' and that they are relying on the eyewitness to be able to solve the crime, means that there is tremendous pressure on the eyewitness to produce *something*. This may be much greater in real-life situations than in the laboratory.

Furthermore, as Wagstaff has emphasized (1982, 1993), many factors are present in the hypnotic interview, quite apart from the hypnotic induction procedure, that may enhance eyewitnesses' performance compared to the standard police interviews that have been described by Fisher, Geiselman and Raymond (1987) and George (1990). For example, witnesses may be more likely to relate their ordeals to a professional psychologist or clinician than a police officer (i.e. because of the interviewer's experience of working as a clinician rather than as a hypnotist); repeated testing; systematic relaxation; establishment of rapport; conducting the interview in a quiet area, which is uninterrupted; techniques to provide memory retrieval cues, such as role-playing and picture drawing; recalling in different orders and reinstating context; and encouraging the adoption of low (Erdelyi, 1992; Hibbard & Worring, 1981; Wagstaff, 1982, 1989, 1993).

Indeed, many of the factors identified by Wagstaff (1982) as being often present in forensic hypnotic interviews, but independent of the hypnotic induction procedure, are very similar to the elements in the 'enhanced' cognitive interview (Fisher, Geiselman & Amador, 1989). Because of this similarity, some have argued that these kinds of non-hypnotic procedures may be just as likely to lead to memory distortion as hypnotic techniques (Perry & Nogrady, 1985), and even that subjects under going those procedures may accidentally fall into a hypnotic trance (Hilgard, 1984; Ofshe, 1992). In support of the former objection, Spanos, Gwynn, Comer, Baltruweit and de Groh (1989) found that a guided imagery procedure produced as much distorted reporting as a hypnotic condition, but the distortion was reversed during cross examination to the levels of controls. On the other hand, Wagstaff, Traverse and Milner (1982) found that by itself, a guided memory procedure did not result in more memory errors, although adding hypnosis to it increased errors without increasing the amount of accurate information. However, only one study has systematically compared hypnosis and the cognitive interview (Geiselman, Fisher MacKinnon & Holland, 1985). In this, both hypnosis and the cognitive interview enhanced recall more than a standard interview (control group), but they did not differ from each other. On the basis of this Geiselman et al., conclude that any positive effects of hypnosis in forensic contexts result from what hypnosis shares in common with the cognitive interview. Neither hypnosis nor the cognitive interview were associated with an increase in false alarms; however, hardly any leading questions were asked, and no confidence measures were taken.

5.7 How permanent are hypnotic memory distortions

As has been mentioned, to non-state cognitive behavioral theorists, the fact that memory distortion sometimes occurs with hypnosis is not really surprising. Hypnotic subjects might make things up, or express spurious confidence, simply because of the additional demands and expectations placed on them by the hypnotic context; i.e. because of the context they may feel socially obliged or pressured to report more information, or believe they are in some special hypermnesic state. Also the fact that highly susceptible subjects may sometimes show more distorted reports in both hypnotic and non-hypnotic situations may occur because such subjects are generally more prone to respond to social pressures (Wagstaff, 1981a, 1986a, 1991a; Spanos et al., 1989). Whether or not studies produce more distortions with hypnosis will depend therefore on the particular social demands and expectations, and individual characteristics of the subjects, present in those studies. But if these are the processes involved, non-state theorists would tend to predict that, when found, hypnotically induced inaccuracies will often reflect reporting biases rather than genuine irreversible memory distortions. A number of recent studies suggest that the former may be the case; indeed hypnotically created pseudomemories can be significantly reversed if subjects are told, for example, that a 'hidden-part' of them can describe their 'real' memories, or are cross-examined under oath, or are given a financial incentive for accurate reporting (Murrey, Cross & Whipple, 1992; Spanos & McLean, 1983; Spanos, Gwynn et al., 1989; Spanos, Quigley et al., 1991). Moreover, Wagstaff and Frost (in press), found that hypnotically created pseudomemories could be reversed if subjects were previously given an opportunity to 'own up' to role-enacting.

5.8 Conclusions concerning hypnotic interviewing

At the beginning of this chapter two criteria were outlined by which hypnosis might be evaluated; 1) Does the technique increase reliable recall? And 2), does the technique produce testimony in which confidence is related to accuracy? To what extent does hypnosis satisfy these criteria?

Hypnosis appears to have a number of drawbacks that appear to limit its forensic usefulness; it may, on occasion increase the amount of correct information that eyewitnesses recall but this is likely to be accompanied by increases in incorrect information also, and an increase the amount of inaccurate information as a proportion of recall.

With regards to C-A relationships hypnosis appears to decrease C-A relationships by increasing confidence in inaccurate responses. However, the situations in which C-A relationships have been evaluated have been essentially biased in order to produce situations that were most likely to produce distorted responses (such as leading questions and misleading post event information), it is plausible that in more neutral situations C-A relationships would be higher, yet such a relationship remains to be investigated.

Furthermore, there is a discrepancy between experimental findings and those reported in the field; field work tends to suggest that hypnosis is a useful interview procedure while experimental work does not. Although there appear to be number of reasons that may explain this (for instance, hypnosis may share elements with the cognitive interview), little work has systematically addressed these issues.

As was pointed out earlier in this chapter, one of the main reasons why hypnosis may 'look' impressive in the field as an interview technique, is that much of the evidence produced by the witness may not be verifiable. However, this problem may not be limited to hypnosis. It is to this issue that we now turn.

CHAPTER 6

THE INFLUENCE OF VERIFICATION ON EYEWITNESS TESTIMONY

6.1 Verification in the field experiments

Many researchers in the field of eyewitness testimony have criticised the lack of ecological validity of eyewitness experiments (see for example, Lindsay & Wells, 1983, Malpass & Devine, 1980; Memon, Holley, Milne, Koehnken & Bull, 1994; Yuille & Cutshall, 1986). Further, police officers and legal professionals are often critical of the extent to which eyewitness testimony experiments can be generalised to the real world. Because of this it would be appropriate to evaluate factors that influence eyewitness performance in real crime situations. However, two main factors limit our ability to use real eyewitnesses.

1. Information that eyewitnesses provide can rarely be verified. In many cases it would not be possible to check if information that a witness provided with a given technique was correct.

2. Novel techniques cannot be used, as an inappropriate technique may adversely influence the quality of eyewitness information that is produced and

therefore may impede the ability of the police to 'solve' a crime or bring criminal charges.

One way of checking eyewitnesses' information is to compare eyewitnesses' descriptions of criminals that they witnessed committing crimes with the actual appearance of criminals convicted of those crimes. However, this has the drawback that the convicted individuals may be innocent (Hollin, 1989; Radkin, 1974; Yuille & Cutshall, 1986). Furthermore, the convictions are likely to have been based on eyewitness accounts in which case eyewitnesses may simply corroborate their own, erroneous testimony. Even if an individual has confessed to having committed the crime this may not necessarily be a sufficient safeguard that the individual is guilty as it appears that there is quite a high rate of false confessions (see, for example, Gudjonsson, 1992).

Problems of verification of eyewitness reports have been encountered in almost all field studies (see for example, Christianson & Hubinette, 1994; Fisher, Geiselman & Amador, 1989; George, 1990) and are a feature of almost all police investigations. In many investigations eyewitnesses' answers cannot be verified.

A paradigm has successfully circumvented these problems is that used by Yuille and Cutshall (1986) in Canada. In this study witnesses had been involved in a shooting incident that occurred outside a gun shop. A thief had entered the gun shop, tied up the proprietor, and stolen some money and a number of guns. The store owner freed himself picked up a revolver and went outside to note the thief's number plate. The thief however, had not yet got into his car and in a face to face encounter on the street, separated by approximately two metres, the thief fired two shots at the shop owner. After a slight pause the shop owner discharged all six shots from his revolver. The thief was killed while the shop owner recovered from his injuries. Witnesses viewed the incident from various vantage points in the street, from nearby buildings or from passing cars, and they witnessed various aspects of the incident either prior to and including the actual shooting or after the shots were fired.

This incident was chosen for further analysis for a number of reasons. 1) there were sufficient witnesses to allow comparisons between witnesses, 2) because the thief was killed and the weapons and stolen money were found by the police, a great deal of forensic evidence was available to verify the accuracy of eyewitness's testimony, 3) the death of the thief closed the police file, allowing the research to proceed without interference in the legal process.

The findings of this study were to some extent different from those of more usual laboratory studies. Witnesses were very accurate in their accounts and there was little change in their recall or their accuracy over five months. The eyewitnesses resisted leading questions, and their stress levels at the time of the event appeared to have no negative effect on subsequent memory. This would seem to indicate that criticisms of the ecological realism of laboratory based experiments may be in some part justified.

Although the above method is a good paradigm with which to assess eyewitness performance, it can rarely be used because such situations occur so infrequently (although see Cutshall & Yuille, 1989); in most cases the police are unable to check many of the answers that eyewitnesses provide. However, as yet little if any research has actually been conducted to systematically assess the effects of verifiability on eyewitness testimony as a variable in its own right, especially when memory facilitation techniques are used.

6.2 Verification in real-life situations

As has been outlined previously, the actual part that eyewitnesses play in criminal proceedings varies greatly from case to case. For example, if a burglar is apprehended by police officers as he runs away from a house that he has just burgled, admits the offence and has also left fingerprints at the crime scene, then there is already enough evidence to convict the burglar of the offence, so it would be unnecessary for the police to conduct in-depth interviews with eyewitnesses.

However, in other cases an eyewitness may be the only individual that witnessed a crime and may also be the only source of evidence. By the very nature of the way in which crimes are investigated, if little information is available from other sources, then greater efforts may be made to obtain as much information as possible from an eyewitness. Thus, if considerable pressure is applied to an eyewitness to recall more information, it is likely to be in cases where very little of what the eyewitness recalls can be verified. This in itself may make the eyewitness feel that there is considerable pressure to recall more information. If the eyewitness is aware that many, or indeed all of the answers that he/she provides cannot be verified, he/she may lower his/her criteria for report, confabulate, and/or express false confidence in an attempt to fulfil the demands of the experimenter.

6.3 Verification and the hypnotic interview

To reiterate from the last chapter, the pressure that eyewitnesses feel to produce more information may be exacerbated by the use of a 'special' memory facilitation interview technique such as 'hypnosis'. In addition, hypnosis is only likely to be used as a last resort, if little information is available from other sources; i.e. in the very situations when the police are least likely to be able to verify answers.

Eyewitnesses that have consented to undergo a hypnotic interview may find themselves under both implicit and explicit pressure, to produce more information. The fact that the police feel that the crime is sufficiently severe to warrant the use of hypnosis means that the witness may be very anxious to help the police solve the crime. Because the police are using hypnosis implies that the technique is effective and that they believe the witness's recall will be enhanced so the witness may feel under pressure to produce some enhancements. It is also clear to the eyewitness that the police have gone to considerable time and trouble to get a hypno-investigator with the expectation that he/she will be able to enhance the witness's memory. The hypno-investigator goes to great lengths to establish rapport and hypnotise the witness with the clear expectation that more information will be produced. In addition, the instructions given under hypnosis may contain explicit suggestions that memory will be enhanced. Thus, a social situation is created which may place great pressure on eyewitnesses to be a 'good' witness and this means produce a great deal of information and express high confidence in that information. Again, if the eyewitness is aware that many, or indeed all of the answers that he/she provides cannot be verified, he/she may lower his/her criteria for report, confabulate, and express false confidence. Indeed, it could be argued that with all the social factors involved, it would be very difficult for an eyewitness undergoing a hypnotic interview to say that they could not remember any additional details.

It was noted earlier that there is a discrepancy in opinion between investigators that have used hypnosis in the field and laboratory findings. Field investigators have tended to extol the use of hypnosis as a memory enhancement procedure in Police investigations (see for example, Haward & Ashworth, 1980; Hibbard & Worring, 1981; Kleinhauz et al., 1977; Reiser, 1980), while laboratory experiments have shown less positive findings (c.f. chapter five). Apart from the factors previously identified, it is possible that this discrepancy may be explained by subjects exaggerating their memory performance in real-life situations because of the pressure that is applied to them to recall more information while they are aware that their answers cannot be checked. On the other hand, whilst hypnotic subjects in laboratory studies may also be under similar pressure to exhibit hypnotic memory enhancement, the fact that they know that there answers may be verified might produce more limited confabulation and false confidence effects.

6.4 Verification and the cognitive interview

The pressure that eyewitnesses' feel to recall information may also be increased by the cognitive interview. For example, when the mnemonics are given to the witness it is clear that they are there to serve a purpose and that purpose is to improve the eyewitnesses' memory. Thus, the cognitive interview combined with the knowledge many, or indeed all of the answers that he/she provides cannot be verified may have the potential to cause the witness to lower his/her criteria for report, confabulate, and/or express false confidence. The fact that the cognitive interview is a 'special' technique, and is considered time-consuming to perform, means that it may be used in similar circumstances to hypnosis, that is to say in the investigation of important crimes where there are no other leads thus where eyewitnesses are aware that they cannot be verified.

6.5 Experimental investigations of verification effects

The only work directly addressing the issue of verifiability in reported memory is the unpublished work at Liverpool University by Tippett (1993). Tippet compared a hypnotic group and control group on their memory for a list of 20 words. These subjects were divided so that half were in verified groups and half in unverified groups.

After presentation of the first word list subjects were given either a two and a half minute delay (control) or a hypnotic induction procedure that took approximately 10 to 15 minutes (hypnosis). Subjects were then given a second list of 50 words and told that all of the words that they had initially been given were present. Subjects in the unverified group simply had to count up the number of words that they could recognise. Thus, the specific words that they had identified could not be verified. In the verified group subjects were given the same instructions as the unverified group but were also required to tick off the words that they could identify, thus they were aware that their answers could be verified.

Subjects in the unverified group stated that they could remember significantly more words from the lists than the verified group. There was no difference between the hypnosis group and the control group in terms of the number of words that they could accurately identify.

Unfortunately, there are two main flaws in this experiment that make interpreting the results difficult. Firstly, the amount of effort that the subjects made may have been different between the unverified and verified groups. The verified group had to actually identify which words they could remember, then write down how many they could remember. The unverified group may not have paid as much attention to each word on the list as they but merely estimated how many words they should have been able to remember. Thus the differences between the groups may have been due to different recall-strategies being used by the subjects in each group. Secondly, different delays between stimuli presentation and testing were used for the hypnosis and control groups. This means that the results could have been confounded by a greater memory trace decay in the hypnosis group compared to the control group. However, this study does indicate that verification may effect subjects' memory.

6.6 Conclusion

In crimes where there are no other leads, considerable pressure may be applied to eyewitnesses to produce more information, this combined with the knowledge that answers cannot be verified may lead eyewitnesses to lower their criteria for report, confabulate, and express false confidence. These problems, if found, may be exacerbated by interview techniques such as hypnosis or the cognitive interview that carry with them explicit or implicit expectations that the eyewitness's recall will be enhanced.

Furthermore, if eyewitnesses do lower their criteria for report, confabulate, and express false confidence if they are aware that answers they provide cannot be verified this may have particularly negative impact on the way in which an investigation or trial is conducted. Because when there is little other information investigating officers or jurors have no alternative but to base their decisions on eyewitness information.

As no research has, to the writer's knowledge, addressed the issue of knowledge that an answer cannot be verified influence eyewitness testimony, and whether this is a particular problem for the cognitive interview, it would appear that such an investigation would be timely.

CHAPTER 7

INTRODUCTION TO THE EXPERIMENTAL WORK

On the basis of the literature review presented so far a number of factors have emerged that form the focus of the present research programme.

7.1 Verification

Chapter six indicated that in some cases an eyewitness may be the only individual who witnessed a crime and may also be the only initial source of evidence. In such cases greater efforts may be made to obtain as much information as possible from an eyewitness. Thus, if considerable pressure is applied to an eyewitness to recall more information, it is likely to be in cases where very little of what the eyewitness recalls can be verified. This in itself may make the eyewitness feel that there is considerable pressure to recall more information. If the eyewitness is aware that many, or indeed all of the answers that he/she provides cannot be verified, he/she may lower his/her criteria for report, confabulate, and express false confidence.

In addition, the pressure on an eyewitness to produce information may be exacerbated by the use of instructions to the eyewitness to motivate him/her to produce more information, especially when used in the context of 'special' memory facilitation interview techniques such as 'hypnosis' or the 'cognitive interview'. These memory facilitation techniques may contain explicit and/or implicit suggestions that memory will be enhanced, creating a social situation which may place great pressure on eyewitnesses to recall extra information. Again, if the eyewitness is aware that many, or indeed all of the answers that he/she provides cannot be verified, he/she may lower his/her criteria for report, confabulate, and express false confidence. As a consequence it would also seem appropriate to investigate the combined effects of interview techniques and verification on memory performance.

7.2 Confidence-Accuracy relationships

In such cases where an eyewitness's testimony cannot be verified with physical forensic information or from other eyewitness accounts, investigating officers must rely on other measures, such as the confidence that an eyewitness expresses in his/her account, to determine accuracy. Thus, if an eyewitness says that he/she is 'absolutely certain' that an assailant had black hair then more emphasis will be placed on that information in a future investigation than if the eyewitness says that he/she was simply 'guessing' that the assailant's hair colour was black. It is plausible that eyewitnesses' confidence in information that they provide, both correct and incorrect, may be mediated by the knowledge that answers cannot be verified, by interview instructions and techniques, or by both.

Previous investigations into C-A relationships have shown little relationship between confidence and accuracy especially when hypnosis is used. However, the procedures that have previously been adopted appear to have been designed to show low C-A relationships. For example, questions of homogenous difficulty have not been used, and many C-A relationships have been investigated after misleading post-event information and/or in response to misleading questions. As a consequence a more neutral approach to measuring C-A relationships would appear warranted.

7.3 Interview techniques

Interview technique may interact with both verification and confidence-accuracy relationships. For example, a hypnotic interview technique may create implicit and explicit demands for subjects to perform well (i.e. to recall a great deal of information and to be confident in the accuracy of that information). These demands may decrease C-A relationships and encourage subjects who are aware that their answers cannot be verified to exaggerate their ability to remember a target event.

In addition, it is possible that an interview technique such as the cognitive interview, may create similar demands and effects to those that have been reported with hypnosis. However, little research has actually been conducted specifically comparing the hypnosis with the cognitive interview in these respects. Therefore, a comparison of a hypnotic interview with a cognitive interview in terms of C-A relationships and verifiability would seem timely.

7.4 Aims of the empirical research

The above considerations give rise to three questions which are addressed in the present thesis:

1) Does the knowledge that answers cannot be verified alter eyewitnesses' responses?

2) Is there a relationship between an eyewitness's confidence and their accuracy?

3) Are 1 and 2 mediated by the interview instructions or memory facilitation techniques?

PART 2

EMPIRICAL RESEARCH

OVERVIEW

Part two describes nine experiments designed to investigate some effects of stimulus verification, interview instructions and confidence on eyewitness memory. It is divided into three parts.

Chapters eight to 14 describe the experimental programme devised to compare performance on a face-recognition task between subjects who were aware that their answers could be checked (a verified group) with subjects that were aware that their answers could not be checked (an unverified group), using a variety of interview instructions.

Chapters 15 to 17 describe the experimental programme devised to investigate the relationship between subjects' confidence and accuracy in response to questions about a video presentation.

Chapter 18 describes an investigation into the effects of hypnosis, the cognitive interview and a control procedure on subjects' performance on a face-recognition task with verifiable and unverifiable answers (verified and unverified groups), and the relationship between subjects' confidence and their accuracy in response to questions about a video presentation.

CHAPTER 8

EXPERIMENT 1: AN EXPERIMENTAL INVESTIGATION OF THE EFFECTS OF VERIFICATION ON A FACE RECOGNITION TASK.

8.1 Introduction

As noted in part one, researchers interested in eyewitness memory have tried many ways of making laboratory research realistic. However, even in situations that may appear realistic, it is usually made clear to subjects that they are taking part in a psychology experiment and hence any answers they produce can be checked against what the experimenter knows to have happened (e.g., George, 1990; Loftus, 1979; Maass & Kohnken, 1989; Yuille & Cutshall, 1984). The aim of the first experiments, therefore, was to devise a paradigm to investigate the possible effects of this factor. On the basis of the issues discussed in Part one it might be expected that unverifiability would result in increases in subjects' reports of how much they can recall compared to a verified group.

A number of possible tasks were considered, such as recall of real life unverifiable events; however, in the end a face-recognition task based on photographs was chosen as most appropriate. This task was considered to be most advantageous, not only because it has obvious relevance to eyewitness testimony in practice, but because of simplicity, repeatability and controllability. Thus, to investigate if knowledge that an answer cannot be verified influences eyewitnesses' performance, Experiment 1 compared the performance of a verified group with an unverified group on a face recognition task.

For the verified group it was also possible to evaluate subjects' actual performance in terms of the number of correct identifications and the number of incorrect identifications that they made. Thus, by also employing an estimate measure, it was possible to determine if there was a relationship between the number of photographs that subjects estimated that they could correctly identify, and the number of correct identifications made; given the literature on confidence-accuracy relationships in eyewitness identification one would not necessarily predict a significant relationship between these measures, however, we might assume the direction to be positive (see, Bothwell, Deffenbacher, & Brigham, 1987; Deffenbacher, 1980; Fruzzetti, Tolland, Teller & Loftus, 1992; Wells & Murray, 1984).

8.1.1 Hypotheses

The main predictions therefore were as follows.

1) Subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified.

2) Subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimate of how many photographs they could identify than subjects whose answers could be verified.

3) Within the verified group there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications they made.

8.2 Method

8.2.1 Subjects

Subjects were 22 females and nine males (age 17-43, mean= 21.70; SD=6.18), recruited from undergraduates at the Psychology Department of Liverpool University or from prospective Psychology students visiting the Department for an open day. First year undergraduate subjects were awarded one credit point for their participation. Subjects were randomly assigned to groups which were either 'verified' (*N*=15) or 'unverified' (*N*=16).

8.2.2 Materials and Procedure

Subjects in both groups were presented with a 'pack' of fifty black and white photographs (75 X 50mm) of adult faces in an unmarked envelope. They were required to shuffle the photographs. Subjects were then instructed to choose 20 photographs at random from the pack and were asked to look at these for a timed period of two minutes.

The most important part of the procedure was to try to ensure that subjects in the 'unverified' group were aware that their answers could not be verified. To achieve this, after the timed period, the 'unverified' group were told to place the photographs that they had chosen back with the photographs from the original 'pack' and to shuffle all the cards together thoroughly, supposedly for subsequent subjects to use. Subjects then placed their shuffled 'packs' in identical plain envelopes which were placed by the subjects themselves in a box passed around the room.

In contrast, after the timed period, the 'verified' group placed the photographs that they had chosen separately into an envelope and were told explicitly by the experimenter that this was so that their answers to any subsequent questions could be verified later.

Subjects in both groups were then given a five minute reading filler activity of two neutral magazine articles after which they were tested on their recognition of the photographs that they had chosen using an identification sheet (300 X 440 mm) that they were told contained all 50 photographs from which they had chosen 20 and given a questionnaire to complete (Questionnaire 1, displayed in Appendix 8.2).

The identification sheet contained all fifty of the stimuli photographs (53 X 35 mm) in five rows of ten. In the bottom left corner of each photograph was a letter and a number for subjects to use for identification purposes. The questionnaire contained a number of filler questions to disguise the experimental hypothesis. The central questions of importance required subjects to 'attempt to identify as many of the twenty photographs that you chose as you can' and later to estimate how many photographs they thought they had accurately identified and to rate their confidence in that answer on a seven point Likert scale ranging from 'not at all confident' (1) to 'very confident' (7).

8.3 Results

The results are divided into two parts, 1) verified/unverified comparisons, and 2) the actual performance, in terms of correct and incorrect identifications, of the verified group.

8.3.1 Verified/unverified comparisons.

The verified and unverified groups were compared in terms of 1) the estimated number of photographs that subjects said they could accurately identify, and 2) the confidence expressed in the accuracy of that estimate. The means and SDs for these analyses are displayed in Table 8.3.1.

No significant difference was found between the verified group and the unverified group with respect to subjects' estimates of the number of photographs which they said they could accurately identify, $F(1,29)=0.13$, $p>.72$, in the confidence that subjects expressed in the accuracy of their estimates of the number of photographs that they could accurately identify,

$F(1,29)=0.95$, $p>.34$[1], or in the number of identifications that they made, $F(1,29)=0.34$, $p>.55$.

Table 8.3.1 Means and SDs of the verified/unverified group comparisons.

measure	verified group N=15	unverified group N=16
estimated no.	14.33 (3.12)	14.75 (3.34)
conf in estimated no.	4.60 (1.50)	5.13 (1.50)
no. of identifications	19.33 (1.35)	18.88 (2.75)

Note: Standard deviations in brackets

Pearson's linear correlations were calculated between subjects' estimates of the number of photographs they thought they could accurately identify, the confidence shown in these estimates and the number of identifications that were made for the verified and unverified groups independently. It can be noted that

1 This analysis used a parametric test on data obtained with a Likert scale. It could be suggested that this data is unsuitable for parametric tests as Likert scales do not produce linear, continuous, data. However, it has been argued that the assumptions often cited as necessary for the use of parametric tests (continuous data, normal distribution, and similar variance) are overly restrictive in practice. In addition, parametric tests are remarkably unaffected by violations of these assumptions (e.g. Howell, 1992). As a consequence, parametric tests are used on Likert-scale data throughout this thesis.

in the present thesis much of the correlational analysis was exploratory, though later a number of the hypotheses regarding correlations can legitimately be viewed as one-tailed. Thus, despite the problems of committing a Type 1 error with a large number of correlations, so as not to miss something that might be of importance, especially with small samples, a two-tailed alpha level of 0.05 was used as the criterion for significance in all experiments.

For the verified group none of these correlations was significant. These results are displayed in Table 8.3.2. along with the correlations for the other variables. When the unverified group alone was considered, there were significant linear correlations between subject's estimates of the number of photographs that they could accurately identify and their confidence shown in the accuracy of that estimate (r=.75) and, also, between subjects' estimates of the number of photographs that they could accurately identify and the number of identifications that they made (r=.57). These correlations are displayed in Table 8.3.3.

Table 8.3.2 Correlations (r) between performance variables for the verified group, N=15.

Measure	1	2	3	4	5	6
1. estimated no.	-	.38	-.05	.40	-.50	.50
2. conf. in 1.		-	-.25	.12	-.34	.30
3. no. of identifications made			-	.52*	.22	-.06
4. no. correct identifications				-	-.72**	.82***
5. no. incorrect responses					-	-.99***
6. accuracy rate						-

Note. $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$.

Table 8.3.3 Correlations (*r*) between performance variables for the unverified group, N=16.

Measure	1	2	3
1. estimated no.	-	.75***	.57*
2. conf. in 1.		-	.33
3. no. of identifications			-

Note. *$p<.05$, **$p<.01$, ***$p<.001$.

8.3.2 Actual performance of verified group

Further analysis, considering actual performance, was possible for subjects in the verified group with regard to number of correct identifications, number of incorrect identifications and accuracy rate (see, Perfect, Watson & Wagstaff, 1993).

Subjects in the verified group made an average of 15.13 (SD=1.88) correct responses and 4.20 (SD=1.66) incorrect responses. The mean accuracy rate (no. correct identifications / no. correct identifications + no. incorrect identifications) was .78 (SD=.08).

There was no significant difference between subjects' estimates of the number of accurate identifications that they could make and the actual number of correct identifications that they made, $F(1,14)=1.03$, $p>.31$ (see Table, 8.3.1).

However, significant correlations were found between the following: the total number of identifications made and the number of correct identifications ($r=.52$); the number of correct identifications and the number of incorrect identifications ($r=-.72$); the number of correct identifications and the accuracy rate ($r=.82$), and the number of incorrect identifications and the accuracy ($r=-.99$). The full matrix is displayed in Table 8.3.2.

8.4 Discussion

8.4.1 Verified/unverified group comparisons

The two main initial hypotheses were: 1) subjects whose answers could not be verified would estimate that they could accurately identify significantly more photographs than subjects whose answers could be verified, and 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimate of the number of photographs that they could accurately identify than subjects whose answers could be verified. Neither of these hypotheses was supported. This might suggest that verifiability actually has no influence on eyewitness judgments. However, there may be a number of other reasons why verification may not have had significant effects in this first experiment.

Firstly, the absence of differences between the two groups' performance may be due to a lack of pressure on the subjects to recall information. If the pressure on subjects to recall more information was increased, for example by the administration of motivating instructions, it is plausible that differences may become apparent between verified and unverified subjects (Wagstaff, 1981).

Secondly, subjects in the unverified group may have felt that there was little reason to exaggerate their estimates of how well they could perform on the face recognition task, because they felt that they could perform well. However, if subjects felt that they were not able to perform well on the face-recognition task unverifiablity may lead subjects to exaggerate their ability to perform to produce more socially desirable responses. The tendency for subjects to alter responses so that they are socially desirable has been widely shown in the social psychological literature (e.g., Asch, 1957; Orne, 1961; Schmidt, Duncan, Taveres, Polanczyk, Pellanda, and Zimmer, 1993; Wagstaff, 1981).

Two correlations were significant for the unverified group, but not the verified group, these were subjects' estimates of the number of photographs that they could accurately identify and their confidence in the accuracy of that answer; and subjects' estimates of the number of photographs that they could identify and the number of identifications that they made. This indicates that subjects who estimated that they could accurately identify more photographs also expressed higher confidence in the accuracy of that estimate and made more identifications. Given that there were no significant differences between the groups in terms of these variables the reason for the differences in the size of the correlations between groups is not obvious. Before speculating, therefore, it was considered advisable to wait until these effects could be replicated.

8.4.2 Actual performance of verified group

A subsidiary hypothesis was that, within the verified group, there would be a positive relationship between subjects' estimates of how many photographs they thought they could accurately identify and the number of correct identifications that they made. This hypothesis was not supported; the correlation between the two measures was not significant. Nevertheless, overall, subjects' estimates of how many accurate identifications they could make was not significantly different from the number of correct identifications that subjects made.

Further analysis showed a significant positive correlation between the total number of identifications made and the number of correct identifications. This does not appear to be because subjects were making more identifications *per se* and so by chance more correct identifications, as there was no corresponding correlation between the number of identifications made and the number of incorrect identifications. Indeed, while the number of correct identifications was significantly correlated with accuracy rate, the number of incorrect identifications was negatively correlated with accuracy rate. Thus, subjects who made more identifications were more likely to make more correct identifications and, proportionately, less incorrect identifications; i.e. there was no evidence for a response bias effect here.

In sum, therefore, the results of this experiment indicated no clear effects for verifiability; in fact there were no obvious response bias effects.

CHAPTER 9

EXPERIMENT 2: AN EXPERIMENTAL INVESTIGATION OF THE EFFECTS OF VERIFICATION, MOTIVATING INSTRUCTIONS AND TRYING AGAIN ON A FACE RECOGNITION TASK.

9.1 Introduction

The two main hypotheses of Experiment 1 were 1) subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified; and 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimates of the number of photographs that they could accurately identify than subjects whose answers could be verified. Neither of these hypotheses was supported.

It was suggested, however, that if the face-recognition task were more difficult, and subjects perceive they were performing badly, subjects in an unverified group might exaggerate their performance compared to a verified group.

Furthermore, if pressure were exerted on subjects to perform better, through motivating instructions, this pressure combined with the knowledge that answers cannot be verified might also lead an unverified group to exaggerate their performance compared to a verified group.

In view of these considerations, the influence of increased task difficulty and motivating instructions on verified and unverified groups was investigated in Experiment 2. Essentially the same procedure was employed so it was also possible to look at subjects' estimates of how many photographs that they could accurately identify and the number of correct identifications they made, under the new conditions. The response sheets were modified, however, so that it was possible to gain an estimate of subjects' confidence in their responses in terms of what they would be prepared to swear to in court.

9.1.1 Hypotheses

The following predictions were again made.

1) Subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified.

2) Subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimates of how many photographs that they could accurately identify than subjects whose answers could be verified.

3) Within the verified group there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications they made.

Plus another,

4) Subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

9.2 Method

9.2.1 Subjects

Subjects were 24 females and 16 males (age 17-59, mean= 20.40; SD=5.98) recruited from undergraduates at the Psychology Department of Liverpool University and from prospective Psychology students visiting the Department for an open day. First year undergraduate subjects were awarded one credit point for their participation. Subjects were randomly assigned to groups which were either 'verified' (N=19) or 'unverified' (N=21).

9.2.2 Materials and Procedure

A similar procedure to Experiment 1 was used. Subjects in both verified and unverified groups were presented with fifty black and white photographs of adult faces which they were required to shuffle. In Experiment 1, subjects chose 20 photographs which they viewed for two minutes; to make the task more difficult, in this experiment, subjects were instructed to choose 25 photographs at random from the pack and were asked to look at these for 30 seconds.

Again, the most important part of the procedure was to try to ensure that subjects in the 'unverified' group were aware that their answers could not be verified. As for Experiment 1, after the timed period, the 'unverified' group were told to place the photographs that they had chosen back with the photographs from the original 'pack' and to shuffle all the cards together thoroughly, supposedly for subsequent subjects to use. Subjects then placed their shuffled 'packs' in identical plain envelopes which were placed by the subjects themselves in a box passed around the room.

In contrast, after the timed period, the 'verified' group placed the photographs that they had chosen separately into an envelope and were told explicitly by the experimenter that this was so that their answers to any subsequent questions could be verified later.

Subjects in both groups were then given a five minute reading filler activity of two neutral magazine articles after which they were tested on their recognition of the photographs that they had chosen using an identification sheet that they were told contained all 50 photographs from which they had chosen 25 and a questionnaire (Questionnaire 1b, displayed in Appendix 9.2.1). The questionnaire was identical to Questionnaire 1 used for Experiment 1 with two exceptions: 1) the question requiring subjects to identify the photographs that they had chosen was removed; and 2) an additional, final, question asked subjects 'if you were asked to testify in court, how many of the twenty-five photographs that you saw, would you be absolutely certain that you saw before?'

Subjects in the verified group were also given an answer sheet (Answer Sheet 1, displayed in Appendix 9.2.2) which they were instructed not to complete until they were asked to do so. The answer sheet required subjects to identify the photographs which they had seen. Five dotted lines were given on which subjects could provide identifications. As subjects in the verified group had been presented with the answer sheet they were made aware that at some point they would have to identify the photographs. The unverified group did not receive a sheet at this stage.

When completed, the questionnaires were collected. Subjects were then given another copy of the questionnaire (1b) and the following instruction:

> Although you have just completed a questionnaire, I would like you to repeat this task. It is very important for this experiment that you try as hard as you can to remember the faces that you were presented with. Please try very hard to remember as much as possible; this is very important.

Subjects in the unverified group were then told that when they had completed the second questionnaire the experiment was over. This ensured that they were aware that there would not be required to identify any photographs that they had seen. Subjects in the verified group were asked to fill in the answer sheet after they had completed the second questionnaire.

It can be noted that in the design of this study no account was taken of whether the numbers of correct and incorrect identifications (as measured in the verified group) were affected by the motivating instructions; i.e. measures of accuracy were only taken **after** the motivating instructions. As the effects of motivating instructions on accuracy per se was not a main point of interest at this stage, this was considered advisable, because if unverified subjects had been asked to actually make identifications, they might have thought they could be checked. On the other hand, if only the verified group had been required to make identifications before the motivating instructions, this might have produced a confounding effect (e.g. unlike the unverified group they might have been subject to intervening cognitive retrieval activities that might have affected their performance).

9.3 Results

The results are divided into two parts 1) verified/unverified comparisons, and 2) actual performance, in terms of correct and incorrect identifications, of the verified group.

9.3.1 Verified/unverified group comparisons.

The performance of verified/unverified groups was again compared in terms of 1) subjects' estimates of the number of photographs that they could accurately identify, and 2) the confidence expressed in the accuracy of that estimate, with the addition of 3), the number of identifications that they would be prepared to testify in court to have seen before. The means and SDs for these analyses are displayed in Table 9.3.1.

The above variables were first analyzed with 2 X 2 ANOVAs with repeated measures on the second factor (verified/unverified X before/after motivating instructions). The results are as follows.

No significant difference was found between the verified and unverified groups with respect to subjects' estimates of the number of photographs that they could accurately identify, $F(1,37)=2.32$ $p>.14$. There was also no significant difference before/after motivating instructions, $F(1,37)=1.17$, $p>.29$ and no significant interaction between verified/unverified group and before/after motivating instructions $F(1,36)=0.02$ $p>.89$.

However, subjects in the unverified group expressed significantly greater confidence in the accuracy of their estimate of how many photographs they could accurately identify than the verified group, $F(1,37)=5.26$, $p<.028$. There was no significant difference before/after motivating instructions, $F(1,37)=1.17$, $p>.29$ and no significant interaction between verified/unverified group and before/after motivating instructions $F(1,37)=0.68$ $p>.42$, in terms of confidence.

No significant differences were found in the number of identifications that subjects stated that they would identify in court with respect to verified/unverified group $F(1,37)=2.23$ $p>.15$. However, for both groups combined, the number of identifications that subjects said they would identify in court did increase after motivating instructions, $F(1,37)=14.82$, $p<.0005$. No significant interaction was found between verified/unverified group and before/after motivating instructions $F(1,37)=0.35$ $p>.56$.

Table 9.3.1. Means and SDs of the verified/unverified group comparisons.

measure	before motivating instructions		after motivating instructions	
	verified N=19	unverified N=21	verified N=19	unverified N=21
estimated no.	12.92 (5.46)	14.81 (4.48)	13.0 (5.59)	15.24 (4.68)
conf. in estimated no.	3.80 (1.14)	4.70 (1.21)	4.04 (1.48)	4.47 (1.12)
identify in court	8.15 (5.00)	10.86 (4.52)	9.41 (5.74)	11.82 (4.66)

Note: Standard deviations in brackets

Pearson's correlations were then calculated for the verified and unverified group independently, before and after motivating instructions, between subjects' estimates of how many photographs they could accurately identify, the confidence shown in the accuracy of those estimates and the number of identifications that subjects said they would identify in court. These correlations are shown in Tables 9.3.2 and 9.3.3.

The majority of significant correlations were significant for both the verified and the unverified groups and these were as follows: subjects' estimates of how many photographs they could accurately identify before and after motivating instructions ($r=.92$ and $r=.79$, respectively); the number of identifications that subjects said they would make in court before motivating instructions ($r=.93$ and $r=.94$ respectively) and after motivating instructions; ($r=.73$ and $r=.81$, respectively); subjects' estimates of how many photographs they could accurately identify and the number of photographs that they said that they would identify in court after motivating instructions ($r=.84$ and $r=.75$, respectively); subjects' estimates of how many photographs they could identify before motivating instructions and the number of identifications that they were prepared to make in court after motivating instructions ($r=.77$ and, $r=.73$ respectively); and there were significant correlations between subjects' estimates of how many photographs they could identify after motivating instructions and the number of identifications that they were prepared to make in court before motivating instructions ($r=.76$ and $r=.81$, respectively).

Two relationships were significant for only one of the groups. The confidence that subjects expressed in the accuracy of their estimates of how many photographs they could accurately identify before motivating instructions was significantly correlated with the same measure after motivating instructions ($r=.61$), but only for the verified group. Subjects' confidence in the accuracy of their estimate of the number of photographs that they could accurately identify before motivating instructions was significantly correlated with the number of identifications subjects would be prepared to make in court after motivating instructions ($r=.48$), but only for the unverified group.

Table 9.3.2. Correlations (r) between performance variables for the verified group, (N = 19)

Measure	1	2	3	4	5	6	7	8	9	10
1. estimated no.	-	.44	$.73^{***}$	$.92^{***}$	.31	$.77^{***}$	$.65^{**}$	$.81^{***}$	.05	.18
2. conf. in 1.		-	.38	.41	$.61^{**}$	.27	-.12	.17	-.45	.43
3. identify in court			-	$.76^{***}$	.13	$.93^{***}$	$.56^{*}$	$.76^{***}$	-.10	.33
4. estimated no.				-	.19	$.84^{***}$	$.67^{**}$	$.84^{***}$	.06	.21
5. conf. in 4.					-	.09	-.06	.06	-.17	.11
6. identify in court						-	$.59^{**}$	$.80^{***}$	-.07	.34
7. no. of identifications made							-	$.82^{***}$	$.67^{**}$	-.35
8. no. correct identifications								-	.12	-.23
9. no. incorrect identifications									-	$-.88^{***}$
10. accuracy rate										-

Note:- $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.
Items 1, 2 and 3 are before motivating instructions. Items 4-10 are after motivating instructions.

Table 9.3.3. Correlations (*r*) between performance variables for the unverified group, $N=20$

Measure	1	2	3	4	5	6
1. estimated no.	-	.31	$.81^{***}$	$.79^{***}$	.21	$.73^{***}$
2. conf. in 1.		-	.46	.40	.21	$.48^{*}$
3. identify in court			-	$.81^{***}$	.18	$.94^{***}$
4. estimated no.				-	-.11	$.75^{***}$
5. conf. in 4.					-	.26
6. identify in court						-

Note:-
$^{*}p<.05$; $^{**}p<.01$; $^{***}p<.001$.
Items 1, 2 and 3 are before motivating instructions. Items 4, 5 and 6 are after motivating instructions.

9.3.2 Actual performance of verified group

Further analysis, considering actual performance, was possible for the verified group with regard to number of correct identifications, number of incorrect identifications and accuracy rate.

Subjects in the verified group made an average of 12.73 (SD=4.42) correct identifications and 4.32 (SD=3.70) incorrect identifications. The accuracy rate was .76 (SD=.15).

A one-way, repeated measures ANOVA with three levels was conducted to compare subjects' estimates of how many accurate identifications they could make both 1) before and 2) after motivating instructions, with 3) the actual number of correct identifications that subjects made. No significant main effect was found, $F(2,34)=.120$, $p>.87$. However, a similar ANOVA comparing the number of identifications that subjects stated that they would identify in court both before and after motivating instructions with the number of correct identifications made showed a significant effect $F(2,34)=21.77$, $p<.001$. Follow-up F-tests for simple effects[2] showed that all three groups were different from one another, all ($p<.05$). Thus the number of identifications that subjects stated they would make in court, both before and after motivating instructions, was significantly less than the number of correct identifications that they made. Subjects increased the number of identifications that they said they would make in court after receiving motivating instructions.

A 2 X 2 ANOVA with repeated measures on both factors (subjects' estimates of the number of photographs they could accurately identify/the number of identifications subjects stated that they would identify in court X before/after motivating instructions) indicated that subjects made significantly greater estimates of the number of photographs that they could accurately identify than the number that they said they could identify in court, $F(1,17)=37.76$ $p<.0001$. There was no significant effect of motivating

---

2

There is a debate within the statistics literature whether it is appropriate to conduct post hoc tests such as Tukey and Scheffe on repeated measures and mixed between/within subjects designs. One way of avoiding the issue is to conduct F comparisons for simple effects as described by Howell (1992), so this was the procedure adopted in this thesis.

instructions, $F(1,17)=3.19$ $p>.09$, and no significant interaction, $F(1,17)=3.26$ $p>.09$.

Significant correlations were found between the following variables: subjects' estimates of how many photographs they could identify both before and after motivating instructions and the number of identifications that they made ($r=.65$ and $r=.67$, respectively); subjects' estimates of how many photographs they could identify both before and after motivating instructions and the number of correct identifications that they made ($r=.81$ and $r=.84$ respectively); the number of identifications that subjects stated they would identify in court both before and after motivating instructions and the number of identifications that they made ($r=.56$ and $r=.59$, respectively); the number of identifications that subjects stated they would identify in court both before and after motivating instructions and the number of correct identifications that they made ($r=.76$ and $r=.80$, respectively). In addition, the number of identifications that subjects made was significantly correlated with the number of correct identifications that they made ($r=.82$). The number of correct identifications was significantly correlated with the number of incorrect identifications ($r=.67$). The number of correct identifications was also correlated negatively with the accuracy rate ($r=-.88$).

9.4 Discussion

9.4.1 Verified/unverified group comparisons

The initial hypotheses relating to verified/unverified comparisons were: 1) subjects whose answers could not be verified would estimate that they could accurately identify significantly more photographs than subjects whose answers could be verified; 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy the estimate of the number of photographs that they could accurately identify than subjects whose answers could be verified; and 3) subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

Of these hypotheses, only the second was supported, that is to say, subjects in the unverified group expressed greater confidence in the accuracy of their estimates of how many photographs they could accurately identify than the verified group. The significant effect of verifiability on this measure, which was not found in Experiment 1, thus may be due to the increased task difficulty of this experiment compared to that of Experiment 1.

No significant interactions were found between the verified and unverified groups and before/after motivating instructions. Therefore, the idea that increasing the motivation of subjects to recall more information would lead subjects in the unverified group to exaggerate their performance was not supported. However, it is plausible that motivating instructions may have different effects on verified and unverified groups in different situations to those that were created in this experiment. For instance, one factor that may have limited the effects of the motivating instructions on the unverified group, was that subjects were required to fill in the questionnaire, then they were given motivating instructions and asked to repeat the same questionnaire. This meant that initial responses could be compared with final responses. Subjects may have reasoned that if they dramatically changed their responses after the motivating instructions they would appear 'fickle', unreliable, or as if they had not been trying hard to begin with. Hence, while the motivating instructions may have increased the pressure on subjects to recall there may have also been considerable pressure on subjects *not* to alter their performance (this kind of 'face saving' effect has been identified in the hypnosis literature, see Wagstaff, 1981, 1986).

Nevertheless, motivating instructions significantly increased the number of identifications that subjects were prepared to identify in court. It would be expected, given the serious consequences associated with court testimony, that this measure would be particularly resistant to change, unless, of course subjects were able to identify more on the second occasion after motivating subjects. However, it is possible that subjects were initially very conservative on this measure; as will be noted shortly, for the subjects on whom accuracy measures were possible, even after the motivating instructions, the number of identifications that they would make in court was still significantly less than the number of identifications subjects could correctly identify. This idea will be considered further in the actual performance (i.e. verified group) section where the number of identifications that subjects would be prepared to identify in court can be compared with the number of correct identifications that they made.

The correlations between performance measures for the verified group and the unverified group were broadly similar. Subjects who made high estimates of the number of photographs that they could identify appeared to do so consistently, whether the measure was their estimates of how many photographs they could accurately identify, the number they would identify in court and regardless of whether this was before/after motivating instructions.

However, some correlations were significant for the verified group but not the unverified group and vice versa. The confidence that subjects expressed in the accuracy of their estimates before motivating instructions was significantly correlated with the same measure after motivating instructions for the verified group only, perhaps indicating that confidence in these estimates was more stable and less exaggerated in the verified group. The confidence that subjects expressed in the accuracy of their estimates before motivating instructions was significantly correlated with the number of identifications subjects would be prepared to make in court after motivating instructions was only significant for the unverified group. The reason for this disparity is not clear, but given the sizes of the correlations (.48 for the unverified group, and .27 for the verified group), and the large overall number of correlations, perhaps not too much should be read into this difference at this stage.

9.4.2 Actual performance of verified group

A subsidiary hypothesis was that, for the verified group, there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications that they made. This hypotheses was supported, there was a significant correlation between these estimates both before and after motivating instructions and the number of correct identifications. It can be noted that this correlation was not significant for Experiment 1, although it was also positive. One possible reason for this is that in experiment 2 the increase in task difficulty increased the variation in subjects' performance, which made differences in subjects performance more apparent.

The number of identifications that subjects stated that they would identify in court both before and after motivating instructions was significantly correlated with the number of correct identifications they made; again this supports the hypothesis that subjects who believed that they could identify more correct photographs actually could do so. Although a crude measure, this would suggest a positive relationship between confidence and accuracy.

When subjects' estimates of how many accurate identifications they could make before and after motivating instructions and the actual number of correct identifications they made, were compared, there was no significant difference between subjects' estimates of how many accurate identifications they could make and the actual number of correct identifications made. Therefore, subjects' estimates of how many photographs they could identify appeared to be an accurate reflection of the number of correct identifications that they made.

However, when the number of identifications that subjects stated that they would identify in court before and after motivating instructions were compared with the number of correct identifications, all three groups were different from one another. Although subjects increased the number of photographs that they stated that they would identify in court after motivating instructions, this estimate was still significantly less than the number of photographs that they could actually identify. Thus, although the number of photographs that subjects stated that they would identify in court was increased by motivating instructions, it was still a conservative estimate, less than the actual number of identifications that subjects could make. This suggests that, despite motivating instructions, subjects might adopt a fairly strict, conservative, criterion for report if they are required to testify in court.

However, there were significant positive correlations between the total number of identifications made, and both the number of correct **and** incorrect identifications made. This was reflected in a negative relationship between the number of identifications made and accuracy. Nevertheless, subjects who made more correct identifications did not make more incorrect identifications (the correlation was only .12). This would suggest an individual difference effect whereby some subjects just make a large number of incorrect identifications, whereas others stick to a more conservative, but accurate strategy of making a fewer number of identifications concentrating on accuracy. It can be noted, however, that as a proportion of the total number of identifications, there were fewer incorrect (25%) than correct responses (75%), suggesting that as the number of identifications increases more correct responses result than incorrect responses.

In sum, there was little support for the view that the effects of verifiability on reports would be influenced by motivating instructions; however, verifiability did appear to increase overall confidence in estimates of accurate identifications. Nevertheless, the other results suggest that, on the whole, subjects' estimates of what the could identify and their actual accuracy were quite reasonable, erring on the side of caution rather than overconfidence.

CHAPTER 10

EXPERIMENT 3: AN EXPERIMENTAL INVESTIGATION OF THE EFFECTS OF VERIFICATION, AND MOTIVATING INSTRUCTIONS ON A SINGLE-TRIAL FACE RECOGNITION TASK.

10.1 Introduction

No significant interactions were found between the verified and unverified groups and before/after motivating instructions in Experiment 2. This suggests that motivating instructions did not have different effects on the verified and unverified groups. However, one factor that may have limited the effects of the motivating instructions on the unverified group was that subjects were required to complete the questionnaire task before and after the motivating instructions. This meant that subjects' initial responses, before motivating instructions, could be compared with responses made after motivating instructions. As has been pointed out in the previous discussion, subjects may have reasoned that if they changed their responses to any great extent after the motivating instructions they would appear 'fickle' or unreliable or as if they had not been trying hard to begin with. Hence, while the motivating instructions may have increased the pressure on subjects to recall there may have also been considerable pressure on subjects *not* to change their performance.

As a consequence, Experiment 3 sought to test the assumption that subjects may be inclined to exaggerate their performance in an unverified group compared to a verified group if motivational instructions are administered *and* there are no baseline performance-levels against which subjects can be compared.

10.1.1 Hypotheses

The following predictions were again made.

1) Subjects whose answers could not be verified would estimate that they could recognise significantly more photographs than subjects whose answers could be verified.

2) Subjects whose answers could not be verified would express significantly greater confidence in the accuracy of the estimate of the number of photographs that they could identify than subjects whose answers could be verified.

3) Subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

4) Within the verified group there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications they made.

10.2 Method

10.2.1 Subjects

Subjects were 22 females and 10 males aged between 17 and 59 recruited from undergraduates at the Psychology Department of Liverpool University or from prospective Psychology students visiting the Department for an open day. First year undergraduate subjects were awarded one credit point for their participation. Subjects were randomly assigned to groups which were either 'verified' ($N=16$) or 'unverified' ($N=16$).

10.2.2 Materials and Procedure

The procedure used for Experiment 3 was similar to the procedure used in Experiment 2. Subjects were presented with stimuli, filler tasks, identification sheets and questionnaires in the same manner as for Experiment 2. The verified group were also given the Answer Sheet, again as for Experiment 2.

However, before undertaking the Questionnaire task all subjects were given the following motivating instructions:

> It is very important for this experiment that you try as hard as you can to remember the faces that you were presented with.

Please try very hard to remember as much as possible, this is very important.

Subjects were not required to repeat the questionnaire or identification task as was required in Experiment 2.

10.3 Results

The results are divided into two parts 1) verified/unverified comparisons, and 2) actual performance, in terms of accuracy, of the verified group.

10.3.1 Verified/unverified comparisons.

Performance of verified/unverified groups was again compared in terms of 1) the estimates of the number of photographs that subjects said they could accurately identify, 2) the confidence expressed in the accuracy of that answer, and 3) the number of identifications that subjects stated that they would make in court. The means and SDs for these analyses are displayed in Table 10.3.1.

The above variables were first analyzed using F tests (verified vs unverified). The results were as follows.

No significant difference was found between the verified group and the unverified group with respect to subjects' estimates of the number of photographs which they could accurately identify, $F(1,31)=0.12$, $p>.73$,. No significant difference was found either between the verified group and the unverified group on the confidence shown in the number of photographs that they could accurately identify, $F(1,31)=3.58$, $p>.07$. And no significant difference was found between the verified group and the unverified group with respect to the number of identifications that they stated that they would make in court, $F(1,30)=0.01$, $p>.92$.

Table 10.3.1. Means and SDs of the verified/unverified group comparisons.

measure	verified group N=16	unverified group N=16
estimated no.	12.19 (4.87)	11.63 (4.24)
conf. in estimated no.	4.31 (0.70)	5.13 (0.9)
identify in court	8.75 (4.46)	8.60 (3.83)

Note: Standard deviations in brackets

Pearson's correlations were performed on subjects' estimates of the number of photographs that they could accurately identify, the confidence shown in these estimates and the number of identifications that subjects would identify in court. The correlations for the verified and unverified groups are shown in Tables 10.4.2., and 10.4.3., respectively. In the both the verified and unverified groups subjects' estimates of how many photographs they could identify and the number of photographs that subjects stated that they would identify in court were significantly correlated ($r=.77$ and $r=.82$, respectively).

Table 10.4.2 Correlations (r) between factors for the verified group.

Measure	1	2	3	4	5	6	7
1. estimated no.	-	-.17	$.77^{***}$	$.52^{*}$	.45	.29	-.11
2. conf. in estimated no.		-	-.21	-.12	-.06	-.15	.10
3. identify in court			-	.49	.41	.30	-.09
4. no. of identifications made				-	$.90^{***}$	$.51^{*}$	.06
5. no. correct identifications					-	.09	.46
6. no. incorrect responses						-	$-.77^{***}$
7. accuracy rate							-

Note. $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$.

Table 10.4.3 Correlations (*r*) between factors for the unverified group.

Measure	1	2	3
1. estimated no.	-	-.16	.82***
2. conf. in estimated no.		-	-.49
3. identify in court			-

Note. *$p<.05$; **$p<.01$; ***$p<.001$.

10.3.2 Actual performance of verified group

As before, further analysis, considering actual performance, was possible for the verified group. Subjects in the verified group made an average of 13.31 (SD=5.21) correct responses and 4.63 (SD=2.66) incorrect responses. The accuracy rate was .74 (SD=.14).

A one-way repeated measures ANOVA between subjects' estimates of how many photographs they could accurately identify (M= 12.19, SD= 4.87), of how many they would identify in court (M= 8.75, SD= 4.46) and the number of correct identifications that they made was significant, $F(2,30)=18.42$ $p<.0006$. Follow-up F tests revealed that there was no significant difference between subjects' estimates of how many photographs they could identify and the number of photographs that they correctly identified, $F(1,15)=72$ $p>.41$. However, subjects stated that they would identify significantly fewer photographs in court than the number of correct identifications that they made, $(1,15)=11.95$ $p<.0035$. Subjects' estimates of how many accurate identifications they could make were also significantly greater than the number of identifications that they said they would identify in court, $F(1,15)=18.42$ $p<.0006$.

Significant correlations were found between subjects' estimates of how many accurate identifications they had made and the number of identifications that they made ($r=.52$), the number of identifications that subjects made and the number that they correctly identified ($r=.90$) and the number of incorrect identifications ($r=.51$). The number of incorrect identifications was significantly negatively correlated with accuracy rate ($r=-.77$).

10.4 Discussion

10.4.1 Verified/unverified comparisons

The three initial hypotheses were: 1) subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified; 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of the estimate of the number of photographs that they could identify than subjects whose answers could be verified; and 3) subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

None of these hypotheses was supported. Thus, it appears that using motivating instructions without subjects being required to complete an initial questionnaire before motivating instructions does not appear to lead subjects in the unverified group to exaggerate their performance. Two explanations for this are plausible.

Firstly, as noted in Experiment 1, subjects may have felt that because it was an experiment on eyewitness testimony it was 'wrong' to exaggerate their performance. That is to say, the demand characteristics inherent in the experimental situation were such that subjects did not feel that it was appropriate to exaggerate their performance. Secondly, it is possible that, although motivating instructions were used in Experiment 3, these instructions were not motivating enough. Perhaps, if stronger motivating instructions were used differences between the groups might become apparent.

When correlations for the verified and unverified groups were compared one correlation was significant in both groups; subjects' estimates of how many photographs they could identify and the number of photographs that subjects said they would identify in court were significantly correlated.

10.4.2 Actual performance of verified group

A subsidiary hypothesis was that there would be a positive relationship between subjects' estimates of how many photographs they thought they could accurately identify, and the number of correct identifications they made. The correlation between subjects' estimates of how many photographs they could accurately identify and the number of correct identifications was positive, but not significant (.45). However, when compared using an ANOVA, subjects' estimates of how many photographs they could accurately identify were not significantly different from the number of photographs they correctly identified. Therefore, it would appear that there was some relationship between subjects' estimates of how many photographs they can accurately identify and the number of correct identifications that they made, but this was weaker than that shown in Experiment 2.

With regards to the number of photographs that subjects stated that they would identify in court, this was significantly smaller than subjects' estimates of how many accurate identifications they could make and the number of correct identifications they did make. This replicates the findings of Experiment 2. The number of identifications that subjects were prepared to make in court appeared to be a very conservative estimate of the number of correct identifications they actually. The correlation between their confidence about identifying in court, and correct identifications was again positive (.41), but unlike in Experiment 2, it was not significant.

There was a significant correlation between subjects' estimates of how many photographs they could accurately identify and the number of identifications that subjects attempted. However, as in Experiment 2, the number of identifications that subjects attempted was significantly correlated with both the number of correct **and** incorrect identifications. But, as before, as a proportion of the total number of identifications, the number of correct answers (74%), was greater than the number of incorrect answers (26%).

Taken together, these results endorse those of Experiment 2, in that verifiability had a negligible effect on responses; moreover, subjects were fairly accurate at estimating their accuracy, yet erred on the side of conservatism when asked about their confidence about making identifications in court.

CHAPTER 11

EXPERIMENT 4: AN EXPERIMENTAL INVESTIGATION OF THE EFFECTS OF VERIFICATION, LEADING-MOTIVATING INSTRUCTIONS AND TRYING AGAIN ON A FACE RECOGNITION TASK.

11.1 Introduction

Motivating instructions were used in Experiment 2 and Experiment 3. The absence of significant before/after interactions between verified/unverified groups for Experiment 2, and absence of differences between verified/unverified groups after motivating instructions for Experiment 3, may suggest that motivating instructions do not produce differences between verified/unverified groups.

However, as mentioned, one possible explanation for the absence of significant differences between the verified and unverified groups might be that the motivating instructions were not strong enough. However, if subjects are given motivating instructions and leading instructions that indicate very obviously that more information is expected, this may lead the unverified group to exaggerate their performance compared to the verified group. Indeed, one could argue that this expectancy might be more obvious in a real life context.

The aim of this next experiment was to investigate this possibility. The results of Experiment 3 showed no evidence that the before/after design used in

Experiment 2 was somehow detrimental to producing effects, so the latter design was used again.

11.1.1 Hypotheses

Again, the following predictions were made:

1) Subjects whose answers could not be verified would estimate that they could recognise significantly more photographs than subjects whose answers could be verified.

2) Subjects whose answers could not be verified would express significantly greater confidence in the accuracy of the estimate of the number of photographs that they could identify than subjects whose answers could be verified.

3) Subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

4) Within the verified group there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications they made..

11.2 Method

11.2.1 Subjects

Subjects were 24 females and 11 males aged between 17 and 59 recruited from undergraduates at the Psychology Department of Liverpool University or from prospective Psychology students visiting the Department for an open day. First year undergraduate subjects were awarded one credit point for their participation. Equal numbers of subjects were randomly assigned to groups which were either 'verified' (N=18) or 'unverified' (N=17).

11.2.2 Materials and Procedure

The procedure used for Experiment 4 was similar to that of Experiment 2. The presentation of stimuli photographs, filler activity and identification sheets (and answer sheet to the verified group) and completion of the questionnaire were all identical to Experiment 2. However, after subjects had completed the questionnaire they were given the following instructions:

> Although you have just completed a questionnaire, I would like you to repeat this task, only this time trying even harder. It is very important for this experiment that you try as hard as possible to remember more of the faces that you were presented with. Please try harder to remember as many more as possible, this is

very important. Indicate how many more that you can identify by writing how many more you can identify compared to the previous number you identified and putting it to the right of the original. If you wish to change any of the answers that you have given use a cross instead of the previous tick.

Subjects in the unverified group were told that when they had completed the questionnaire task the experiment was over. The verified group were informed that they were also required to complete the answer sheet.

11.3 Results

The results are divided into two parts 1) verified/unverified comparisons, and 2) actual performance, in terms of accuracy, of the verified group.

11.3.1 Verified/unverified group comparisons

The performance of verified/unverified groups was compared in terms of 1) the estimated number of photographs that subjects said they could accurately identify, 2) the confidence expressed in the accuracy of that estimate and, 3) the number of identifications that they would be prepared to testify in court to have seen before. The means and SDs for these analyses are displayed in Table 11.3.1.

The above variables were analyzed with 2 X 2 ANOVAs with repeated measures on the second factor (verified/unverified X before/after motivating instructions). No significant difference was found between verified and unverified groups with respect to subjects' estimates of the number of photographs that they stated they could accurately identify, $F(1,33)=0.16$ $p>.69$. There was also no significant difference before/after leading-motivating instructions, $F(1,33)=0.41$ $p>.53$, and no significant interaction was found between verified/unverified group and before/after leading-motivating instructions $F(1,33)=3.26$ $p>.08$.

No significant difference was found between verified and unverified groups with respect to the confidence they expressed in the accuracy of their estimates of how many photographs they could accurately identify, $F(1,33)=.36$ $p>.56$. There was also no significant difference before/after motivating instructions in subjects' confidence in the accuracy of their estimates of how many photographs they could accurately identify after leading-motivating instructions, $F(1,33)=0.41$ $p>.53$. And no significant interaction was found between verified/unverified group and before/after leading-motivating instructions $F(1,33)=0.87$ $p>.36$.

No significant difference was found in the number of identifications that subjects stated that they would identify in court with respect to verified/unverified group $F(1,32)=0.32$ $p>.57$. However, the number of identifications that subjects said they would identify in court did increase after motivating instructions, $F(1,32)=8.20$ $p<.0073$. No significant interaction was found between verified/unverified group and before/after motivating instructions $F(1,32)=0.86$ $p>.36$.

Table 11.3.1. Means and SDs of the verified/unverified group comparisons.

measure	before leading motivating instructions		after leading motivating instructions	
	verified N=18	unverified N=17	verified N=18	unverified N=17
estimated no.	10.33 (3.05)	10.18 (3.71)	10.72 (3.58)	11.82 (3.99)
conf. in estimated no.	4.00 (1.03)	4.35 (1.06)	4.06 (1.16)	4.06 (0.90)
identify in court	4.67 (3.71)	5.63 (3.05)	5.89 (3.72)	6.25 (3.44)

Note: Standard deviations in brackets

Pearson's correlations were calculated for the verified and unverified group independently, before and after motivating instructions, between subjects' estimates of the number of photographs they thought they could accurately identify, the confidence shown in those estimates and the number of identifications that they said they would identify in court. These correlations are shown in tables 11.3.2 and 11.3.3.

The correlations that were significant for both the verified and the unverified group and were as follows: subjects' estimates of how many photographs they stated that they could accurately identify before/after leading-motivating instructions were significantly correlated (r=.74 and r=.94, respectively); the number of identifications that subjects said they would make in court before/after motivating instructions were significantly correlated (r=.87

and $r=.84$ respectively); subjects' estimates of how many photographs they stated they could identify and the number of photographs that they said that they would identify in court were significantly correlated before motivating instructions ($r=.69$ and $r=.64$, respectively), and after motivating instructions ($r=.64$ and $r=.54$, respectively); and there was a significant correlation between subjects' estimates of how many photographs they could identify after motivating instructions and the number of identifications that they were prepared to make in court before motivating instructions ($r=.57$ and $r=.72$, respectively).

Three correlations were significant for the verified group but not the unverified group. These were as follows. Subjects' estimates of how many photographs they could identify before motivating instructions and the number of identifications that they were prepared to make in court after motivating instructions ($r=.68$). The correlation between subjects' confidence in the accuracy of their estimate before motivating instructions was significantly correlated with the same measure after motivating instructions ($r=.59$), and also the number of identifications that subject would identify in court before motivating instructions ($r=.55$).

Table 11.3.2. Correlations (r) between factors for the verified group, N=18.

Measure	1	2	3	4	5	6	7	8	9	10
1. estimated no.	-	.28	.69*	.74***	.04	.68**	.45	.44	-.13	.10
2. conf. in 1.		-	.55*	.21	.59**	.37	.11	.05	-.19	-.09
3. identify in court			-	.57*	.28	.87***	.42	.27	-.06	-.23
4.estimated no.				-	.29	.54*	.72***	.70**	.12	-.08
5. conf. in 4.					-	.21	.40	.34	.05	-.19
6. identify in court						-	.33	.20	-.09	-.17
7. no. of identifications made							-	.86***	.55*	-.27
8. no. correct identifications								-	.20	.22
9. no. incorrect identifications									-	-.67**
10. accuracy rate										-

Note:- *$p<.05$, **$p<.01$, ***$p<.001$. *Items 1,2 and 3 are before motivating instructions. Items 4-10 are after motivating instructions.*

Table 11.3.3. Correlations (r) between factors for the unverified group, N=17.

Measure	1	2	3	4	5	6
1. estimated no.	-	-.24	.64**	.94***	.00	.49
2. conf. in 1.		-	-.03	-.18	.24	.16
3. identify in court			-	.72***	.00	.84***
4.estimated no.				-	-.05	.64**
5. conf. in 4.					-	.15
6. identify in court						-

Note:-
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.
Items 1,2 and 3 are before motivating instructions. Items 4, 5 and 6 are after motivating instructions.

11.3.2 Actual performance of verified group

Further analysis, considering actual performance, was conducted for the verified group with regard to number of correct identifications, number of incorrect identifications and accuracy rate.

Subjects in the verified group made an average 9.50 (SD=3.54) correct identifications and 3.11 (SD=2.72) incorrect identifications. The accuracy rate was .74 (SD=.16).

A one-way repeated measures ANOVA, with three levels, was conducted comparing subjects' estimate of how many accurate identifications they could make both before and after motivating instructions and the actual number of correct identifications. There was no significant difference between subjects' estimates of how many accurate identifications they could make and the actual amount of accurate identifications made $F(2,34)=1.64$ $p>.21$. However, when a one-way repeated measures ANOVA was conducted that compared the number of identifications subjects stated that they would identify in court before and after leading-motivating instructions and the number of correct identifications, a significant difference was found $F(2,34)=15.50$ $p<.0003$. Follow-up F tests showed that all three groups were different from one another ($p<.05$). The number of identifications that subjects stated they would make in court both before and after motivating instructions was less than the actual number of accurate identifications that they made. Subjects increased the number of identifications that they stated they would make in court after motivating instructions were administered.

A 2 X 2 ANOVA with repeated measures on both factors (subjects' estimates of the number of photographs they could accurately identify/the number of identifications subjects stated that they would identify in court X before/after motivating instructions) indicated that subjects made significantly higher estimates of the number of photographs that they could accurately identify than the number that they said they could identify in court, $F(1,13)=48.04$ $p<.0001$. There was no significant effect of motivating instructions, $F(1,13)=1.22$ $p>.29$, and no significant interaction, $F(1,13)=0.89$ $p>.36$.

Significant correlations were found between the following variables: subjects' estimates of how many photographs they could identify after motivating instructions and the number of identifications that they made (r=.72); subjects' estimates of how many photographs they could identify after motivating instructions and the number of correct identifications that they made (r=.70); the number of identifications that subjects made was significantly correlated with both the number of correct identifications that subjects made (r=.86) and the number of incorrect identifications that subjects made (r=.55). The number of incorrect identifications was also negatively correlated with accuracy rate (r=-.67).

11.4 Discussion

11.4.1 Verified/unverified group comparisons

The three initial hypotheses were: 1) subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified; 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of the estimate of the number of photographs that they could identify than subjects whose answers could be verified; and 3) subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

No significant interactions were found between the verified and unverified groups and before/after motivating instructions. Therefore, there was no support for the prediction that increasing the motivation of subjects to recall more information would lead subjects in the unverified group to exaggerate their performance. It is still possible, however, that, although the face-recognition task was more difficult than that used in Experiment 1, subjects may still have felt they were performing well enough for there to be no need to make-up information.

A significant effect of motivating instructions on the number of identifications that subjects were prepared to identify in court was found; this replicates the findings of Experiment 2 (i.e. there was an increase).

The correlations between performance measures for the verified group and the unverified group were also broadly similar. Subjects who made high estimates of the number of photographs that they could identify appeared to do so consistently, whether the measure was their estimate of how many photographs they could accurately identify, the number they would identify in court and regardless of whether this was before/after motivating instructions. However, some other correlations were only significant for the verified group. For the verified group only, subjects' confidence in the accuracy of their estimates before motivating instructions was significantly correlated with the same measure after motivating instructions. Also the confidence of verified subjects in the accuracy of their estimates of the number of photographs they could accurately identify before motivating instructions was significantly correlated with the number of photographs they would identify in court before motivating instructions. This suggests that subjects who were confident in their estimates were also more likely to say that they would identify more photographs in court, perhaps because the more confident one is, the more one would feel able to testify in court. Perhaps, together, these results indicate that confidence in estimates is more stable and reliable when answers can be verified. However, the latter relationship disappeared with the addition of motivating instructions for both groups, suggesting perhaps that the increase in confidence about testifying in court as a consequence of the motivating instructions upset this stability.

11.4.2 Actual performance of verified group

A subsidiary hypothesis was that there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications that they made. This hypotheses was supported to some extent; there was a significant correlation between these estimates after motivating instructions and the number of correct identifications, although this was not the case for subjects' initial estimates of how many photographs they could accurately identify. The number of identifications that subjects stated that they would identify in court both before and after motivating instructions was significantly correlated with the number of correct identifications that they made, again supporting the hypothesis that subjects who believed that they could identify more correct photographs actually did so.

When subjects' estimates of how many accurate identifications they could make before and after motivating instructions and the actual number of correct identifications subjects made were compared, there was no significant difference between subjects' estimates of how many accurate identifications they could make and the actual number of correct identifications made. Therefore, subjects' estimates of how many photographs they could identify appeared to be related to the number of correct identifications that they made.

As in Experiment 2, the motivating instructions led to an increase in the confidence of making identifications in court, but when the number of identifications that subjects stated that they would identify in court before and after motivating instructions were compared with the number of correct identifications; all three groups were different from one another. Although the number of photographs that subjects stated that they would identify in court increased after motivating instructions, it was still a conservative estimate, less than the actual number of identifications that subjects could make.

As in the two previous experiments, the number of identifications attempted was significantly correlated with both the number of correct identifications and the number of incorrect identifications; however, again, as a proportion of the total responses, the proportion of correct responses (75%) was greater than the proportion of incorrect responses (25%). Indeed, the 75/25 proportion in this respect emerges as very consistent.

On the whole, therefore the results of this experiment are very similar to those of the previous experiments.

CHAPTER 12

EXPERIMENT 5: AN EXPERIMENTAL INVESTIGATION OF THE EFFECTS OF VERIFICATION, AND MOTIVATING INSTRUCTIONS ON A SINGLE-TRIAL FACE RECOGNITION TASK, IN WHICH STIMULI FACES WERE ABSENT ON THE RECOGNITION TASK.

12.1 Introduction

Although an increase in task difficulty in Experiment 2 produced a difference between the verified and unverified conditions, i.e. unverified subjects showed higher confidence in the accuracy of their estimates of how many photographs they could identify, similar differences were not found between verified and unverified groups on similar tasks in Experiment 3 and Experiment 4.

Consequently, in this next experiment task difficulty was increased further, to determine if this would eventually produce more consistent differences in performance between the verified and unverified groups. This was achieved by using one set of photographs of adult faces as the stimulus material, and another different set on the identification sheet. Thus, subjects were required to identify photographs on an identification sheet that were not presented as stimuli. Also extra instructions were given to make it clear to unverified group that their answers could not be verified.

Clearly, because no correct responses were possible, with the exception of a refusal to make responses, the verified groups' actual performance could not be considered.

12.1.1 Hypotheses

The following predictions were made.

1) Subjects whose answers could not be verified would estimate that they could recognise significantly more photographs than subjects whose answers could be verified.

2) Subjects whose answers could not be verified would express significantly greater confidence in the accuracy of the estimate of the number of photographs that they could identify than subjects whose answers could be verified.

3) Subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

12.2 Method

12.2.1 Subjects

Subjects were 30 females and eight males (age 17-43, mean= 18.50; SD=6.18) recruited from prospective Psychology students visiting the Department for an open day. Subjects were randomly assigned to groups which were either 'verified' (N=19) or 'unverified' (N=19).

12.2.2 Materials and Procedure

Subjects were tested in groups as part of their visit to the Department for an open day. Subjects were aware that they had been split into these groups and that part of their tour of the department would include a brief experiment.

The procedure and materials used for Experiment 5 were similar to that for Experiment 2. The presentation of stimulus photographs was identical to that of Experiment 2, with the exception that the unverified group received the following instructions before they attempted to identify the photographs.

> Usually we would ask you to keep the photographs that you chose separately from the pack from which you chose them, so that we can check your answers later on. However, today, because we are testing a number of groups with no breaks in between there is no time for us to check your answers before the next group

arrives. So please shuffle the photographs which you chose with the other left-over photographs and place them back into the envelope.

The verified group were, as in previous experiments asked to put the photographs that they had chosen in the envelope next to their questionnaire and explicitly told that this was so that their answers could be verified.

After the filler activity, subjects in the verified group were given the Answer Sheet although they were instructed not to fill it in until they were told to do so. All subjects were given an identification sheet and the Questionnaire (1b). The identification sheet used in this experiment had the same format identification sheet used in Experiments 1 to 4, but did not actually contain any of the stimulus photographs; instead fifty alternative, similar, photographs, were substituted. Subjects were given the following information before they completed the questionnaire.

The identification sheet contains photographs of all fifty people who were shown on the fifty photographs from which you chose twenty-five earlier. Although they are all present they may be disguised. For example an individuals' hair may have changed in style or colour. Facial hair may have changed, for example, a beard may be present on an individuals' photograph on the identification sheet which was not present on the original photograph.

Subjects in the verified group were told that it was not possible for them to fill in the Answer Sheet due to time restrictions.

12.3 Results

12.3.1 Verified/unverified comparisons

Performance of verified/unverified was compared in terms of 1) the estimates of the number of photographs that subjects said they could accurately identify, 2) the confidence expressed in the accuracy of that answer, and 3) the number of identifications that subjects stated that they would make in court. The means and SDs for these analyses are displayed in Table 12.3.1.

The above variables were analyzed using *F* tests (verified vs unverified). The results were as follows.

The verified group's estimate of how many accurate identifications they could make was significantly higher than the estimate of the unverified group, $F(1,36)=5.06$ $p<.03$. However, no significant difference was found between the verified group and the unverified group on the confidence shown in the number of photographs that they could accurately identify, $F(1,36)=2.11$ $p<.15$. And no significant difference was found between the verified group and the unverified group in the number of identifications that they would make in court, $F(1,30)=0.01$, $p<.92$.

Table 12.3.1. Means and SDs of the verified/unverified group comparisons.

measure	verified group N=19	unverified group N=19
estimated no.	8.58 (4.10)	5.63 (3.98)
conf in estimated no.	3.26 (1.33)	3.84 (1.12)
identify in court	4.16 (4.57)	2.84 (3.83)

Note: Standard deviations in brackets

Pearson's correlations were performed on subjects' estimates of the number of photographs that they could accurately identify, the confidence shown in this estimate and the number of identifications that subjects would identify in court.

For the unverified group subjects' estimate of how many accurate identifications they had made was significantly correlated with the number of identifications that they would make in court (r=.91), while for the verified group the same trend was evident, the correlation between these variables was not significant (r=.46). Furthermore, in both the verified and unverified groups the confidence that subjects expressed in their estimates of how many photographs they could accurately identify and the number of photographs that subjects would identify in court were significantly correlated (r=.63 and r=.46, respectively). See Tables 12.3.2 and 12.3.4.

Table 12.3.2 Correlations (r) between factors for the verified group, N=19.

Measure	1	2	3
1. estimated no.	-	.30	.44
2. conf. in estimated no.		-	.63**
3. identify in court			-

Note. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 12.3.3 Correlations (r) between factors for the unverified group, N=19.

Measure	1	2	3
1. estimated no.	-	.37	.91***
2. conf. in estimated no.		-	.46*
3. identify in court			-

Note. *$p < .05$, **$p < .01$, ***$p < .001$.

12.4 Discussion

12.4.1 Verified/unverified comparisons

The three initial hypotheses were: 1) subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified; 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of the estimate of the number of photographs that they could identify than subjects whose answers could be verified; and 3) subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court.

None of these hypotheses was significant. Indeed, the only significant difference found was in the opposite direction to that which was predicted; i.e. those in the verified group estimated they could recognize more photographs than those in the unverified group. After so many nonsignificant comparisons this could have been a statistical fluke. Alternatively there might be a reasonable explanation that might turn the predictions made earlier on their head. Reference has been made earlier to the demand characteristics inherent in experimental situations (see Kelman, 1967; Orne, 1961; Pierce, 1908, Wagstaff, 1981). That is, in undertaking a 'psychology' experiment, subjects are aware that the experiment has a purpose; and in this instance the experiment concerned eyewitnesses' performance. Furthermore, the experimenter explicitly told the subjects in the unverified group that, although answers can usually be checked, they would not be checked on this occasion. may have tried to help the experimenter. As they were aware that the experimenter could not check their answers but usually could, they may have been especially conservative in order not to mislead the experimenter. The verified group on the other hand, aware that their answers could be checked for accuracy may have been more willing to make errors because any errors that they produced would have been of little consequence (as the experimenter would be aware of them).

Nevertheless, regardless of the interpretation of this finding, clearly, an increase in item difficulty alone does not produce a situation whereby subjects' who are aware that their answers cannot be verified exaggerate their performance compared to a group in which they can be verified.

For the unverified group subjects' estimate of how many accurate identifications they had made was significantly correlated with the number of identifications that they would make in court (because perhaps, for the reason stated above, they had been more careful), while for the verified group there was a trend in the same direction between these variables but the correlation was not significant. However, in both the verified and unverified groups the confidence that subjects expressed in their estimates of how many photographs they could accurately identify and the number of photographs that subjects would identify in court were significantly correlated, indicating some consistency.

CHAPTER 13

PRELIMINARY CONCLUSIONS WITH REGARD TO THE EFFECTS OF STIMULUS VERIFICATION ON EYEWITNESS PERFORMANCE.

13.1 Verified/unverified comparisons

Two main hypotheses were considered for Experiments 1 to 5: 1) that subjects whose answers could not be verified would estimate that they could recognise significantly more photographs than subjects whose answers could be verified, and, 2) that subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimate of the number of photographs that they could identify than subjects whose answers could be verified. In addition a third hypothesis that subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court was considered in Experiments 2 to 5.

No significant differences in the predicted direction were found between verified and unverified groups in terms of subjects' estimates of how many accurate identifications they could make, or the number of identifications that they would make in court. Only two comparisons were significant. In Experiment 2, subjects in the unverified group were more likely to state that they were confident in their estimates of how many photographs they could accurately identify. And, in Experiment 5, unverified subjects were **less** likely to exaggerate their estimates.

Perhaps, as intimated earlier, the demand characteristics inherent in experiments on eyewitness performance of the kind presented here might actually encourage **accurate** reporting, rather than exaggeration. If eyewitnesses in real-life situations construe interview sessions in the same way, i.e. they feel that providing exaggerated information may hamper Police investigations, the present results might indicate that exaggeration through unverification is not a major problem in eyewitness testimony. Nevertheless, although the last experiment presumably exhausted the possibilities for task difficulty, it could still conceivably be argued that in real-life situations, particularly in important cases where there are few leads, eyewitnesses may be placed under considerably more pressure to produce further information, than in the present experiments so far; particularly if explicit memory facilitation techniques are employed, such as hypnosis or cognitive interview procedures. The possible effects of using such procedures will be considered in later experiments.

Broadly speaking, the correlations between performance measures for the verified group and the unverified group were broadly similar. Subjects who made high estimates of the number of photographs that they could identify appeared to do so consistently, whether the measure was their estimate of how many photographs they could accurately identify, the number they would identify in court and regardless of whether this was before/after motivating instructions where they were used. However some discrepancies did occur between the verified and unverified groups, and, on the whole these tended to favour the verified group; i.e. they suggested more consistency and reliability in the estimates of the verified groups.

On the whole, therefore, the five experiments did not produce robust differences or trends in performance between the verified and unverified groups.

13.2 Actual performance of verified group

Broadly speaking, there was considerable similarity across experiments when actual performance was considered. Subjects' estimates of how many photographs they could accurately identify was not significantly different from the number of correct identifications that they could correctly identify. This lends some support to the hypotheses that eyewitnesses can accurately judge their memory performance.

The number of identifications that subjects were prepared to identify in court (where tested) was significantly less than the number of correct identifications that subjects could make and also significantly less than subject's estimates of how many photographs they could accurately identify. For Experiment 2 and 4 the number of identifications subjects would identify in court increased significantly after motivating instructions, however, the number of identifications that subjects would make was still less than their estimates of how many accurate identifications they could make or the number of correct identifications that they did make. Thus, it would appear that where court

testimony is concerned, although motivating instructions influenced judgements, subjects still adopted a conservative criterion for report.
However, with the exception of Experiment 1, subjects who made more correct identifications also made more incorrect identifications, though the proportion was consistently 75/25 in favour of the former.

One of the more reliable findings in the experiments so far was that subjects' estimates of the number of accurate identifications that they could make and the number of identifications made were, on the whole, significantly correlated. Thus there was some indication of a positive relationship between confidence and accuracy. It is to this issue that we now turn.

As mentioned in the introduction, perhaps one of the reasons why confidence-accuracy relationships have rarely been measured in studies of memory facilitation is that researchers have assumed there is no relationship. Thus, before attempting a more rigorous investigation into the possible effects of motivating memory facilitation instructions on confidence and accuracy it is necessary to examine the conditions under which such a relationship is most likely.

CHAPTER 14

EXPERIMENT 6: THE INFLUENCE OF ITEM DIFFICULTY ON THE RELATIONSHIP BETWEEN EYEWITNESS CONFIDENCE AND ACCURACY ON A TWO-CHOICE FORCED RECALL TASK.

14.1 Introduction

As stated in the introduction, although it might appear intuitively obvious that there is a strong positive relationship between eyewitnesses' confidence and their accuracy, much research in this area appears to contradict this assumption. For instance, reviews by, for example, Bothwell, Deffenbacher, & Brigham (1987), Deffenbacher (1980), Fruzzetti, Tolland, Teller & Loftus (1992) and Wells & Murray (1984), suggest that there is either no relationship, or only a small positive relationship between eyewitnesses' confidence and their accuracy (c.f. chapter two). In an attempt to explain these apparently counter-intuitive findings, Smith, Kassin & Ellsworth (1989) suggested that researchers have concentrated on the confidence-accuracy (C-A) relationships 'between-subjects', comparing the accuracy of confident witnesses to less confident witnesses, rather than the relationship within subjects' own statements. In the latter case, an eyewitness may say that he/she is absolutely certain of some things but is not at all certain of others. To assess within-subject and between-subject confidence accuracy relationships Smith et al. (1989) showed subjects a slide presentation followed by a number of two-alternative forced choice questions. They were then required to rate their confidence in each answer on a ten-point scale. The average between-subjects and within-subjects C-A correlations were comparatively low, $r = .14$ and $r = .17$ for between and within subjects measures respectively. Smith et al. concluded that confidence is not a useful indicator of the accuracy of a particular witness or of the accuracy of particular statements made by the same witness.

However, one factor that has yet to be systematically investigated is item difficulty. Typically in work in this area, researchers attempt to select items so as to avoid floor and ceiling effects; i.e. they try to avoid items that are either very easy or very hard to remember. But in real-life forensic investigations some questions that eyewitnesses are asked may clearly be easier to answer than others. It may be the case, therefore, that previous researchers may have chosen unrealistic and overly homogeneous pools of items, thus reducing the variance necessary for high correlations.

Another possibly important factor is that of the relationship between 'absolutely certain' responses and accuracy. This effect may be precluded when 'easy' items are excluded, and, as it is not necessarily related to correlation size, it may often be missed in correlational analysis (Gruneberg and Sykes, 1993). However, regardless of overall C-A accuracy, it could be the case that the relationship between these 'absolutely certain' responses and accuracy remains high.

The purpose of Experiment 6 was to examine these proposals. This experiment was constructed to be similar to that of Smith et al. (1989) in that subjects were required to answer two-alternative forced-choice questions about a film which they had watched. However, in addition, questions were devised to fall into one of two categories, easy or hard.

14.1.1 Hypotheses

There were two main hypotheses.

1) There would be a large, positive relationship between eyewitness confidence and accuracy if item difficulty is varied by including easy and hard questions.

2) If a subject stated that he/she was 'absolutely certain that an answer he/she had provided was correct, the answer would very likely be correct.

14.2 Method

14.2.1 Subjects

Subjects were 51 prospective University students (37 females, 14 males) visiting Liverpool University Psychology Department for an open day (mean age= 19 years, range 17-29, SD= 3.40).

14.2.2 Materials and Procedure

Subjects were tested in groups varying in size from 14 to 21. They were shown a five and a half minute colour film via a video player and television monitor. The film concerned an elderly couple and their doctor in their home.

Following the film subjects were given a five minute filler task, followed by a questionnaire devised to test their recall of the film (Questionnaire 2, displayed in Appendix 14.2). The format was similar to that used by Smith et al. (1989). Twenty questions were used, each followed by a forced choice of two alternative answers. The questions were devised by two experimenters so that there were ten questions in each of two categories; easy or hard. For example, an easy question was 'What sex was the person in the bed?', while a hard question was, 'was there a picture of a zebra or a horse on the wall?' Respondents were required to rate their confidence in each answer on a ten point Likert scale ranging from 'pure guess' (1) to 'absolutely certain'(10).

14.3 Results

The mean correct responses for the two categories were compared (see Table 14.3.1); the result was highly significant, subjects were more likely to answer easy questions correctly than hard questions, $t(50)=20.22$, $p<.0001$.

C-A correlation was calculated for each subject across the 20 questions (easy and hard questions combined)[3]. This resulted in 50 of what Smith et al. (1989) refer to as 'within subjects' correlations (the N's for various calculations vary because in some cases, when all questions were answered correctly or

---

3

In this case a convenentional Pearsons' product-moment correlation could not be used because of the dichotomous nature of correct/incorrect answers. As a consequence point-biserial correlations were used for these analyses. Algerbraically a point-biserial correlation is equal to a Pearsons' product-moment correlation, thus, the figure r will be used throughout this thesis to include point-biserial correlations.

incorrectly it was not possible to use the data). The average of these correlations was $r=.54$, SD=.15. To test whether this correlation was significantly different from zero, the procedure used by Smith et al. (1989) was adopted. The correlation coefficient for each subject was transformed into a z score and the average z score was tested against zero[4]. The result was significant, $t(49)=3.18$, $p<.005$.

Within-subject C-A relationships were also calculated for easy and hard questions independently; i.e. each subject's C-A correlation was calculated for the 10 easy or 10 hard questions then averaged across subjects. These correlations ($r=.50$, SD=.37 and $r=.17$, SD=.32 for the easy and hard questions respectively) were tested against zero again using converted z scores, but neither was significant.

Subjects' average confidence scores were then calculated for the categories of easy questions answered correctly, easy questions answered incorrectly, hard questions answered correctly and hard questions answered incorrectly. These are summarised in Table 15.3.1. A two-way repeated measures ANOVA (2 X 2, question difficulty X correct/incorrect answer) was conducted on these data (only 12 of the 51 subjects could be used for this analysis as 36 subjects answered all of the easy questions correctly, two subjects answered all of the easy questions correctly and all of the hard questions incorrectly, and one subject answered all of the easy and the hard questions correctly).

---

4 r values were converted to Fisher's z scores to give a better approximation of a normal distribution (Howell, 1992), and to allow comparison with the work of Smith et al. who used this procedure.

Results showed that average confidence was significantly higher in correctly answered questions than in incorrectly answered questions, $F(1,11)=43.25$, $p<.0001$, and the average confidence expressed in easy questions was significantly greater than that expressed in hard questions, $F(1,11)=167.81$, $p<.0001$. A significant interaction was also found between question difficulty and correct/incorrect answers $F=(1,11)=10.84$, $p<.01$. Follow up F tests for simple effects showed that all the means were significantly different ($p<.05$) from each other, but whilst the difference between average confidence in correct easy questions was very much greater than the average confidence expressed in easy incorrect answers, the difference between average confidence in hard questions which were answered correctly and incorrectly, although in the same direction, was not so great. It can also be noted that subjects reported greater confidence in their correct responses to easy questions than to hard questions, and that this relationship was maintained for incorrect answers; i.e. subjects reported greater confidence in easy questions which were answered incorrectly than hard questions which were answered incorrectly.

To assess what Smith et al. term 'between subjects effects, first, each subject's average accuracy was correlated with his/her average confidence rating. Overall (easy and hard questions combined) this correlation was $r=.26$, $p<.05$ (N= 51). However, whereas for hard questions alone this correlation was only .21, $p<.10$, for easy questions it was $r=.54$, $p<.0001$.

When the average accuracy rate and average confidence score for each question was correlated for all 20 questions (easy and hard combined) the correlation was $r=.74$, $p<.0005$. For the ten easy questions alone the correlation was still significant, $r=.74$, $p<.025$; but not for the ten hard questions, $r=.12$, $p>.10$. The C-A correlation across-subjects for each of the 20 questions (easy and hard combined) was then calculated. Only 16 correlations could be used because on four of the easy items all subjects were correct. The average of these correlations ($r=.22$, SD=.21) was not significantly different from zero. The between subjects C-A correlation for the six easy questions ($r=.38$, SD=.26), was again higher than that for the 10 hard questions ($r=.14$, SD=.12), but neither was significantly different from zero.

Table 14.3.1. Mean correct and incorrect answers, and confidence ratings for Experiment 6 (maximum values, 10 for easy and hard questions, and 10 for absolute confidence).

Measure	Easy questions	Hard questions	Overall
Correct	9.65 (0.72)	5.18 (1.53)	14.83 (1.84)
Incorrect	0.33 (0.71)	4.80 (1.54)	5.14 (1.85)
Confidence correct	9.02 (0.73) **8.74** **(0.76)**	2.69 (1.29) **3.28** **(1.77)**	6.95 (0.80) **6.01** **(1.02)**
Confidence incorrect	4.88 (2.43) **4.88** **(2.43)**	2.06 (1.21) **1.99** **(1.14)**	2.18 (1.28) **3.42** **(1.73)**

Note. Standard deviations are in brackets, means in bold were used for the analysis of variance.

When 'absolutely certain' responses are considered, out of a total of 1016 overall responses, 321 of the answers given by subjects were rated as being 'absolutely certain' that their answers were correct; and of these 321 answers 319 were correct, an accuracy rate of 99.4%.

14.4 Discussion

The fact that subjects were significantly more likely to answer easy questions correctly than hard questions validated the classification of easy and hard questions.

When C-A correlations for each subject across questions (easy and hard questions combined) was averaged a significant, positive relationship was revealed. However, when easy and hard were considered questions independently these correlations were not significant. These findings suggest that, when a heterogeneous pool of items is used, within-subjects C-A correlations can be higher than has been found previously.

The ANOVA analysis showed that subjects' average confidence was significantly higher in correct answers than in incorrect answers and thus supports the assertion that there is a positive relationship between subjects' confidence and their accuracy. It can also be noted that subjects reported greater confidence in their correct responses to easy questions than to hard questions, and that this relationship was maintained for incorrect answers; i.e. subjects reported greater confidence in easy questions which were answered incorrectly than hard questions which were answered incorrectly.

When between subjects effects were assessed significant correlations were found overall and for easy questions, but not for hard questions alone.

Thus, on the whole, subjects who were generally more confident in their answers were more accurate than less confident subjects, but, this appeared to be mainly due to their responses to easy questions.

When the average accuracy rate and average confidence score for each question were considered a fairly substantial correlation was produced. Thus, the greater the confidence that subjects **overall**, or as a group, expressed in answers to a particular question the greater the likelihood that it would be answered correctly. This relationship was maintained for easy questions alone, but not hard questions alone.

The C-A correlations across-subjects for each question (either combined or easy and hard questions independently) were not significant. Thus, an individual's confidence was not a good predictor of accuracy at the level of individual items; however, it should be borne in mind that 40% of the scores of the easiest items, i.e. those items on which subjects were most likely to be correct and confident, could not used in this analysis. To emphasise the latter point, when subjects were 'absolutely certain' about a response then they were very unlikely to be inaccurate.

CHAPTER 15: EXPERIMENT 7

THE INFLUENCE OF ITEM DIFFICULTY ON THE RELATIONSHIP BETWEEN EYEWITNESS CONFIDENCE AND ACCURACY ON A FORCED FREE-RECALL TASK.

15.1 Introduction

Although the manipulations used in Experiment 6 were relatively successful in raising C-A relationships on most measures, it could still be argued that, as in the experiment by Smith et al. (1989), because many questions could be answered correctly by chance alone, the C-A relationships were artificially lowered. Hence a further experiment was conducted in which this was taken into consideration. In the following Experiment, the design was similar to that of Experiment 6, but to reduce the effects of guessing, rather than using two-alternative forced choice questions, open-ended questions were used. Also, as a simple selection of 'easy' and 'hard' questions might be considered to be over restrictive and unrealistic, questions were selected so as to be divided into three categories of difficulty: easy, medium, and hard.

15.1.1 Hypotheses

Two main hypotheses were tested.

1) There would be a large, positive relationship between eyewitness confidence and accuracy if item difficulty is varied by including easy, medium and hard questions

2) If a subject stated that he/she was 'absolutely certain that an answer he/she had provided was correct, the answer would very likely be correct.

15.2 Method

15.2.1 Subjects

Subjects were 45 undergraduate Psychology students (32 females, 13 males). The mean age was 23 years (range 18-43, SD= 6.10).

15.2.2 Materials and Procedure

Subjects were tested in two groups, one of 17 and one of 28. They were shown a 5 and a half minute black and white video film that concerned the implied murder of a male by a female.

Following the film, subjects were given a 10 minute filler task, followed by a 33 item questionnaire devised to test their recall of the film (Questionnaire 3, displayed in Appendix 15.2.2). The format of the questionnaire was similar to that of Experiment 6. The questions were open-ended, but subjects were required to provide an answer, even if this was only a guess. The questionnaire was devised by two experimenters who agreed on 11 questions in each of three categories of item difficulty; easy, medium or hard. To reduce the influence of being correct by chance, each question was devised such that a range of plausible answers was possible. For example, an easy question was 'what song was the woman singing?', a medium difficulty question was 'what was on the dish next to the television set?', and a hard question was 'what was behind the Tabasco sauce bottle in the kitchen?' After answering each question subjects were required to rate their confidence in their answer on a ten point Likert scale as for Experiment 6.

15.3 Results

A one-way repeated-measures ANOVA showed a significant effect of question difficulty on the number of correct answers $F(2,88)=591.37$, $p<.0001$. Follow up F tests ($p<0.05$) confirmed that these differences were in the appropriate direction; easy questions were more likely to be answered correctly than medium questions, which in turn were more likely to be answered correctly than hard questions. (see Table 15.3.1).

A C-A correlation was calculated for each subject across the 33 questions and these correlations were averaged for the 45 subjects. The average within-subject correlation (easy, medium and hard combined) was $r=.78$, SD=.08, which was significantly different from zero, $t(44)=5.00$, $p<.0005$.

Within-subject C-A relationships were also calculated for easy, medium and hard questions independently; i.e. each subject's C-A correlation was calculated for the 11 easy, medium or hard questions then averaged across subjects. For easy questions the average within-subjects correlation was .76, SD=.21, which was significantly different from zero, $t(23)=2.26$, $p<.025$ (only 24 of the 45 subjects were used in this analysis as 21 subjects answered all of the easy questions correctly). For the medium difficulty questions the average within-subjects correlation was .55, SD=0.23, which was also significantly different from zero, $t(44)=1.82$, $p<.05$. However, for the hard questions the average within-subjects correlation of .34 was not significantly different from zero (only 20 of the 45 subjects could be used in this analysis as 25 subjects answered all of the hard questions incorrectly).

Subjects' average confidence scores were also calculated for six categories of question response: easy questions answered correctly; easy questions answered incorrectly; medium questions answered correctly; medium questions answered incorrectly; hard questions answered correctly; and hard questions answered incorrectly. These data are summarised in Table 15.3.1. A two-way repeated-measures ANOVA (3 X 2, question difficulty X correct\incorrect answer) was conducted on these data (only 12 of the 45 subjects could be used for this analysis because many subjects answered all of the easy questions correctly, all of the hard questions incorrectly or both). There was a significant main effect for question difficulty, $F(2,20)=26.53$, $p<.0001$; follow up F tests for simple effects showed that all means were significantly different from each other. Subjects expressed greater confidence the easier the questions. There was also a significant main effect for confidence in correct/incorrect answers. Average confidence was higher for correctly answered questions than for incorrectly answered questions, $F(1,10)=96.18$, $p<.0001$. The interaction was not significant.

Each subject's average accuracy was correlated with his/her average confidence rating ($N=45$). Overall (easy, medium and hard questions combined) the correlation was $r=.56$, $p<.0001$. For easy questions alone this correlation was $r=.69$, $p<.0001$, for medium questions $r=.45$, $p<.01$, and for hard questions $r=.45$, $p<.01$.

The average accuracy rate and average confidence score for each question was then correlated. For all 23 questions (easy, medium and hard combined) this correlation was $r=.97$, $p<.0001$. For the 11 easy questions the correlation was $r=.96$, $p<.0001$; for the 11 medium difficulty questions, $r=.84$, $p<.005$ and for the 11 hard questions, $r=.29$.

The average C-A correlation across-subjects for each of the 33 questions (easy, medium and hard combined) was $r=.49$ (SD=.35), however, this was not significantly different from zero (only 26 of the 33 questions could be used in this analysis as three of the easy questions were answered correctly by all subjects and four of the hard questions were answered incorrectly by all subjects). Similar correlations were then separately calculated for easy, medium and hard questions. For easy questions the average correlation was $r=.63$, SD=.27, which was significantly different from zero, $t(7)=1.85$, $p<0.05$.

The average correlations for medium questions, .43 (SD= .38), and hard questions, .42 (SD=.38), however, did not differ significantly from zero.

Table 15.3.1. Mean correct and incorrect answers, and confidence ratings for Experiment 7 (maximum values 11 for easy medium and hard questions, and 10 for absolute confidence).

Measure	Item difficulty			
	Easy	Medium	Hard	Overall
Correct	9.93 (1.16)	4.58 (1.98)	0.67 (0.80)	14.78 (2.82)
Incorrect	1.07 (1.16)	6.42 (1.98)	10.29 (0.82)	18.20 (2.82)
Confidence correct	9.00 (0.90) **9.18** **(0.74)**	6.25 (2.05) **7.06** **(1.26)**	3.39 (3.00) **3.88** **(3.30)**	7.90 (0.98) **6.71** **(1.2)**
Confidence incorrect	3.02 (2.43) **4.21** **(2.94)**	2.63 (1.13) **3.30** **(1.60)**	1.41 (0.48) **1.37** **(0.38)**	1.93 (0.62) **3.01** **(1.09)**

Note. Standard deviations are in brackets, means in bold were used for the analysis of variance.

Also, to illustrate again the general importance of including items about which subjects could be 'absolutely certain', it can be noted that, out of a total of 1481 overall responses, 387 of the answers given by subjects were rated as being 'absolutely certain' that their answers were correct. Of these 387 answers 377 were correct, an accuracy rate of 97.4%, again supporting the assertion that subjects who rate their confidence as 'absolutely certain' are very unlikely to be inaccurate.

15.4 Discussion

The experimenter's categorisation of item difficulty was supported by the fact that easy questions were more likely to be answered correctly than medium questions, which in turn were more likely to be answered correctly than hard questions.

When the average within-subject correlation for easy, medium and hard questions combined was considered it was significant, large and positive. The corresponding independent correlations for easy, medium and hard questions revealed significant correlations for the easy and medium questions but not for the hard questions. This appeared to be a consistent theme running through the results; C-A relationships were higher for easy questions than for medium questions which were in turn higher than for hard questions. When the ANOVA of subjects' average confidence scores was considered, subjects showed greater confidence in correct answers than incorrect answers. In addition, subjects expressed greater confidence the easier the questions. The latter finding will shortly be discussed in the next chapter.

Again, on the whole, these results suggest that using questions of varied difficulty, *within-subjects*, may produce higher C-A relationships than have usually found. Subjects are more confident about their correct answers than their incorrect answers, though, in general, less so for difficult items, thus the hypothesis that there is a large, positive relationship between eyewitness confidence and accuracy if item difficulty is varied by including easy, medium and hard questions was supported.

When each subject's average accuracy was correlated with his/her average confidence rating either overall or for easy, medium questions or hard questions these correlations were significant. These results suggest again that, subjects who were more confident about their answers were also more likely to be more accurate, even when answering hard questions.

When correlations between average accuracy rate and average confidence score for each question were considered, significant correlations were found for easy, medium and hard questions combined, easy questions alone and medium questions alone but not for hard questions alone was not significant. Again, this shows that, overall, the greater the confidence that subjects expressed, as a group, in answers to a particular question the greater the likelihood that subjects would answer that question accurately.

However, the average C-A correlation across-subjects for each of the questions was not significant overall, although there was a significant correlation when easy questions were considered independently. Thus to some degree these findings replicate those in Experiment 6; that is, when extreme items were excluded, an individual's degree of confidence in a particular item was generally a poor predictor of accuracy. However, even so, on this occasion, despite the truncated set of items, a significant C-A relationship was still maintained for easy items alone.

Again, to illustrate the general importance of including items about which subjects could be 'absolutely certain', it can be noted that subjects who rate their confidence as 'absolutely certain' are very unlikely to be inaccurate.

CHAPTER 16: PRELIMINARY CONCLUSIONS

PRELIMINARY CONCLUSIONS CONCERNING THE ROLE OF ITEM DIFFICULTY ON CONFIDENCE-ACCURACY RELATIONSHIPS.

The results of the correlational and ANOVA analyses in both experiments support the view that, when questions which vary in difficulty are used, and thereby maximise the probabilities of producing 'absolutely certain' and 'pure guess' responses, confidence accuracy relationships are, in general, considerably higher than have been previously been reported (Kassin et al., 1989; Perfect et al., 1993).

The results also illustrate how C-A relationships may be affected by the way items are tested. The ANOVAs in both experiments indicated that, the easier the questions, the more confident and more accurate subjects were in their responses to them, but, in Experiment 6, when within-subject C-A correlations were calculated independently for different levels of question difficulty, no significant correlations were found. This is not surprising given that the data used in the correlations (which would concern only easy **or** difficult items) would be more homogeneous than the data used in the ANOVA (which would consist of responses to **both** easy and difficult items). Nevertheless, in Experiment 7, there were still significant C-A correlations for questions of easy and medium difficulty but not for hard questions. In accordance with the rationale presented earlier, such discrepancies between the two studies may have been due to the fact that in the Experiment 6 a two alternative forced choice task was used, which would increase the probability of guessing correctly by chance, and thus lower the C-A correlations (see Perfect et al., 1993).

The trend for C-A relationships tend to be higher for easy items was a recurrent theme in the results. This could possibly have happened because of a difference in the kind of errors that occurred for the different question types. For easy questions, the majority of subjects provided correct answers and displayed high levels of confidence in the accuracy of their answers. Therefore, what few errors did occur may have been due to factors such as lapses of attention, or momentary distraction at the encoding stage. Consequently, those subjects who missed some easy stimuli would have expressed low confidence and showed poor accuracy to questions that concerned this information, thus producing a high C-A correlation for easy questions. However, for more difficult questions, even subjects who paid close attention to the film were unable to answer many of the questions correctly, and the few accurate responses they did make, many have resulted from guessing. Such guessing would again lower the C-A relationship for difficult items (Perfect et al., 1993).

When the average accuracy rate and average confidence scores for each question were correlated, and the correlations averaged, the overall correlations were particularly impressive (.74 and .97 for Experiments 6 and 7 respectively). Though in both studies this particular correlation was low and not significant for difficult questions alone. If generalised these results suggest that, unless all the questions are difficult, when a group of witnesses say 'they are confident about X Y and Z, but not A B or C' their general confidence may quite reliably predict their general accuracy. It can be noted that these correlations in particular would have benefitted from the principle of aggregation (Ajzen, 1988), whereby higher and more reliable correlations are produced by correlating aggregate scores from a number of data subsets (here aggregated over subjects), than by performing separate correlations for each subset, and then averaging the correlations.

Consistently least reliable, however, were the averaged C-A relationships for individual questions; thus when individuals' responses to individual items were considered, confident subjects were not significantly more likely to be correct than less confident subjects, unless all items were easy. This was perhaps not unexpected as these correlations would have benefitted least from the principle of aggregation. But even so, this should not be taken to mean that C-A responses of individuals to individual items are necessarily unreliable. These particular correlations were based on a restricted range of items that did not include responses to the most easy and the most difficult questions (they excluded those items on which all subjects were correct or incorrect, because it was not possible to calculate C-A correlations for these particular questions). And balancing this particular set of results, was the finding that when individual subjects were 'absolutely sure' of a response to an individual question, they were very likely to be correct. This is perhaps more important, as in courts of law it is this kind of definitive response that seems to carry the most impact (Brigham & Wolfskeil, 1983; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Thomas, 1988; Fox & Walters, 1986; Lieppe, Manion & Romanczyk, 1992; Lindsay, Wells & O'Connor, 1989; Wells Ferguson & Lindsay, 1981).

Taken as a whole, therefore, the results might suggest that high C-A relationships are most likely to occur when a) the items to be remembered are relatively heterogeneous in terms of difficulty; b) the calculations are performed on aggregate scores, and c) subjects are 'absolutely sure' of their responses. Having established this, we can now look at the possible effects of hypnosis and the cognitive interview on C-A relationships.

CHAPTER 17: EXPERIMENT 8

AN EXPERIMENTAL INVESTIGATION OF THE EFFECTS OF VERIFICATION AND HYPNOSIS ON A FACE RECOGNITION TASK.

17.1 Introduction

As pointed in Chapter 4, those who use hypnosis in the field have tended to argue that hypnosis is an effective forensic interview technique (see for example, Haward and Ashworth, 1980; Hibbard and Worring, 1981; Kleinhauz et al., 1977; Reiser, 1980); however, laboratory experiments tend to indicate that 'hypnosis' does not dramatically enhance recall compared to motivated 'non-hypnotic' procedures. If anything, hypnosis simply serves to create a more lax criterion for report, and increases confidence in inaccurate as well as accurate responses (e.g. Smith, 1983; Wagstaff, 1993, 1989, 1995). One possible reason for this discrepancy may be that in real-life situations frequently eyewitness' responses cannot definitively be checked; thus whilst there may **appear** to be increases in recall with hypnosis, there is rarely any way of objectively assessing the validity of all the items recalled (c.f. Chapter 5).

Assuming that the demand characteristics of the interview situation produce a situation in which subjects are highly motivated to perform well, then given the experimental literature that suggests that hypnosis may lower subjects' criteria for report and produce false confidence, one might expect that if subjects do exaggerate their performance when they are aware that answers that they provide cannot be verified, then they would be more likely to do so when a hypnotic procedure is used. Consequently, a pilot study was conducted to address this issue. This compared the memory performance of two 'hypnotic' groups on the same recognition task as in experiments 2 to 4; one for which the answers could be verified and one which they could not.

To assess the extent to which 'hypnosis' per se might be influential in these processes, subjects were also classified in terms of hypnotic susceptibility. This procedure was based on the that of McGlashan, Evans and Orne (1969), according to the rationale that, 'If a response is due to the presence of hypnosis, then it should be present only in the hypnotizable subjects' (p.230). Thus, as a measure of susceptibility, subjects were required to report their state of hypnotic depth according to the Long Stanford Scale (Tart,1970). This scale was chosen because of its ease of use, although it correlates highly with other measures of hypnotic susceptibility. For instance, Tart (1970) reports various correlations between the LSS and the SHSS:C (Stanford Hypnotic Susceptibility Scale: Form C); the relevant correlations for the procedure as used here are .61 and .58 between the LSS and the SHSS:C behavioural score, and .79 and .63 between the LSS and the SHSS:C experiential score. Tart reports that the correlation between the experiential and behavioural scores of the SHSS:C is only .77; this is fractionally less than the correlation between undeliberated LSS depth reports and the experiential score on the SHSS:C (which is .79). When one also considers that typical correlations between experiential score on the SHSS:C and for example, the SPS2 (Standard Profile Scales), The HGSHS (Harvard Group Scale of Hypnotic Susceptibility) the BSS (Barber Suggestibility Scale) are .72, .59,and .58, respectively (see Bowers, 1983 for a review), then the

LSS would appear to be as valid a measure of hypnotic susceptibility as many other measures in common use.

Also, although the task was essentially the same as in experiments 2 to 4, in view of the previous findings on the relationship between confidence and accuracy, an additional, and more detailed measure of confidence was employed.

17.1.1 Hypotheses

The following predictions were made.

1) Subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified.

2) Subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimates of how many photographs that they could accurately identify than subjects whose answers could be verified.

3) Subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court than subjects whose answers could be verified.

4) Subjects whose answers could not be verified would express greater confidence in the accuracy of their identifications than subjects whose answers could be verified.

5) The above effects, if present, would be more obvious in subjects classified as highly susceptible.

6) Within the verified group, there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications they made. However, if hypnosis increases spurious confidence it is possible that this relationship would be lower in hypnotically susceptible subjects.

17.2 Method

17.2.1 Subjects

A total of 27 subjects, 16 female 11 male (mean age= 21.80, SD=4.60) of various backgrounds participated in the experiment. Equal numbers of subjects were randomly assigned to groups which were either 'verified' or 'unverified'.

17.2.2 Materials and Procedure

The procedure used for Experiment 8 was similar to the procedure used for Experiment 2. However, so that the reader does not have to refer back to Experiment 2 the entire procedure for this experiment is given here.

Subjects in both groups were presented with a 'pack' of fifty photographs of faces which they were required to shuffle. Subjects were then instructed to choose 25 photographs at random from the pack and were asked to look at these for 30 seconds.

The most important part of the procedure was to ensure that subjects in the 'unverified' group were aware that their answers could not be verified. To achieve this, after the timed period, the 'unverified' group were told to place the photographs which they had chosen back with the photographs from the original 'pack' and to shuffle all the cards together thoroughly, supposedly for subsequent subjects to use. Subjects then placed their shuffled 'packs' in identical plain envelopes which were placed by the subjects themselves in a box passed around the room.

In contrast, after the timed period, the 'verified' group placed the photographs which they had chosen separately into an envelope and were told explicitly by the experimenter that this was so that their answers to any subsequent questions could be checked later.

Subjects in both groups were then given a five minute reading filler activity after which they were tested on their recognition of the photographs that they had chosen using an identification sheet which they were told contained all 50 photographs from which they had chosen 25 and a questionnaire

(Questionnaire 1b). The questionnaire contained a number of filler questions to disguise the experimental hypothesis. The central questions of importance required subjects to estimate how many photographs they thought they could identify. Subjects were required to rate their confidence in that answer on a seven point Likert scale ranging from 'not at all confident' (1) to 'very confident' (7). Subjects were also asked how many photographs they would be prepared to identify, with absolute certainty, in court.

Once this task had been completed, subjects were then played a standard 5 minute taped 'hypnotic' induction procedure. This was a slightly modified version of the widely used induction procedure provided by Barber (1969, pp. 251-254) for use with the Barber Suggestibility Scale. Before being given the induction termination instructions (i.e. which commence 'you are going to wake up in a few minutes'), subjects were first required to report their state of hypnotic depth according to the Long Stanford Scale (Tart,1970); this (the LSS) requires subjects to state their depth of hypnosis on a scale from '0 to 10 (for instance, 0 indicates 'awake and alert', 2 means 'lightly hypnotised', 5 means 'quite strongly and deeply hypnotized', 8 means 'very hypnotised', and 10 means 'very deeply hypnotised and will do almost anything'). The LSS was modified to enable subjects to write down their answers (see Wagstaff & Ovenden, 1979). Following this subjects were given four instructions for memory facilitation. These were based on the four mnemonics used by Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian & Prosk (1984) in their 'cognitive interview' procedure. The reason for using these mnemonics more commonly associated with the cognitive interview is that if we are to untangle the effects of hypnosis per se from cognitive interviewing in general it is essential that the memory facilitation instructions remain constant. There is no standard 'hypnotic interview', but the techniques do overlap with those of the cognitive interview (Wagstaff, 1982). These instructions were as follows.

1) Reinstate context. 'What I would like you to do now is think carefully about the photographs that you saw. Think about what you felt and what you thought when you looked at the photographs. Think about how you were feeling at the time, and of your reactions to the photographs. Try to reinstate the context in your mind, of the physical environment in which you saw the photographs, such as the location of objects and people in the room.'

2) Report everything. 'It is known that some witnesses hold back information- because they are not sure about what they can remember or don't know if the information is relevant. However, you must try to report everything that you can'.

3) Recall the events in different orders. 'Most people remember details of an event in a certain order, from beginning to end. However, you should also try to recall the photographs that you saw in a different order- perhaps starting at the end of the period that you viewed the photographs and working backwards. Alternatively, you can start at a photograph you can remember particularly well and work either forwards or backwards from there'.

4) Change perspectives. 'People who witness events sometimes try to remember events from somebody else's perspective, so try to remember what you would have seen if you were a different person viewing the photographs'.

It can be noted that in this experiment the physical context was not actually changed so the efficacy of, in particular, the context reinstatement mnemonic would presumably be negligible in actually enhancing accurate recall through a context reinstatement mechanism. However, it should be emphasized that the principle aim of this study was to assess the effects of the mnemonics on biases in reporting.

Subjects were then instructed as follows.

Please open your eyes, whilst remaining hypnotized, and repeat the questionnaire task and fill out the answer sheet.

The Answer Sheet 1b is displayed in Appendix 17.2.2. The answer sheet consisted of 25 spaces in which to make identifications of the photographs shown on the identification sheet. Subjects were required to state their confidence in each identification on a seven point likert scale identical to that used to measure confidence in Experiment 2. They were instructed as follows.

When you have completed the questionnaire and the answer sheet close your eyes again, whilst remaining hypnotised, and await further instructions.

When all subjects had completed this task and had closed their eyes, some subjects had not made 25 identifications on the answer sheet. Further instructions were given to those subjects to open their eyes, to draw a line under their last identification and carry out the following instruction:

> If you are not at all sure which of the remaining photographs were the ones that you chose, fill in the remaining spaces with photographs which you think were most likely to have been present.

Subjects who had made 25 identifications were instructed to:

> 'remain relaxed with your eyes closed'

On final completion of the answer sheet, subjects were given the induction termination instructions from Barber (1969); i.e. they were 'woken up' by counting from five to one. Subjects were then debriefed and thanked for their participation.

17.3 Results

The results are divided into two parts 1) verified/unverified comparisons, and 2) actual performance, in terms of accuracy and including C-A relationships, of the verified group.

17.3.1 Verified/unverified comparisons.

An F test (verified v's unverified) on hypnotic depth scores revealed no significant difference between the verified and unverified groups in terms of hypnotic susceptibility , $F(1,25)=0.04$ $p > .84$. Means and SDs are displayed in Table 17.3.1. It can be noted that in both cases the mean score lay between 2 and 3, approximating to 'lightly' to 'quite hypnotised' on the LSS. The range was from 0 to 7. Subjects were divided into 'high' or 'low' hypnotisability according to their LSS depth scores; scores equal to or greater than 3 indicating 'high', and scores less then 3 indicating 'low'.

Table 17.3.1. Hypnotisability of subjects with respect to verified/unverified group.

group		
verified N=13	unverified N=14	combined N=27
2.46 (1.30)	2.62 (2.56)	2.56 (2.22)

Note: Standard deviations in brackets.

Subjects' estimated number of photographs that they could accurately identify, the confidence expressed in the accuracy of that answer, and the number of identifications that they would be prepared to testify to in court, were analyzed with 2 X 2 X 2 mixed ANOVAs with repeated measures on the third factor (verified/unverified X high/low hypnotisability X before/after motivating instructions). The means and SDs for these analyses are displayed in Table 17.3.2.

No significant differences were found in the number of photographs that subjects estimated that they could identify with respect to verified/unverified group $F(1,23)=0.12$ $p>.73$, or of hypnotic susceptibility $F(1,23)=2.61$ $p>.12$. There was, however, a significant increase in the number of photographs that subjects estimated that they could identify after hypnosis $F(1,23)=9.19$ $p<.006$. There were no significant interactions between the verified/unverified groups before/after hypnosis, between before/after hypnosis and hypnotic susceptibility, between verified/unverified groups and before/after interview instructions, and between verified/unverified group, before/after interview and hypnotic susceptibility, $F(1,23)=1.75$ $p>.20$, $F(1,23)=1.78$ $p>.20$, $F(1,23)=0.91$ $p>.35$, respectively.

No significant differences were found in the confidence expressed in the accuracy of the number of photographs that subjects estimated that they could identify with respect to verified/unverified group $F(1,23)=4.16$ $p>.05$, or of hypnotic susceptibility $F(1,23)=0.22$ $p>.65$ or before/after hypnosis $F(1,23)=0.73$ $p>.40$. There were no significant interactions between verified/unverified group before/after hypnosis, between before/after hypnosis and hypnotic susceptibility, between verified/unverified groups and before/after interview instructions, and between verified/unverified group, before/after interview and hypnotic susceptibility, $F(1,23)=2.85$ $p>.11$, $F(1,23)=0.41$ $p>.53$, $F(1,23)=2.17$ $p>.53$, respectively.

No significant differences were found in the number of photographs that subjects said that they would identify in court with respect to verified/unverified group $F(1,23)=0.12$ $p>.74$, or of hypnotic susceptibility $F(1,23)=0.61$ $p>.44$. However, there was a significant increase in the number of photographs that subjects stated they would identify in court after hypnosis, $F(1,23)=12.60$ $p<.002$. There were no significant interactions between verified/unverified group before/after hypnosis, between before/after hypnosis and hypnotic susceptibility, between verified/unverified groups and before/after interview instructions, and between verified/unverified group, before/after interview and hypnotic susceptibility, $F(1,23)=3.86$ $p>.06$, $F(1,23)=2.02$ $p>.17$, $F(1,23)=0.15$ $p>.70$, respectively.

Subjects were only required to make identifications after hypnosis; it can be noted, however, that after hypnosis, if subjects had not made 25 identifications of the photographs that they had chosen they were instructed to make more identifications until they had made 25 identifications; this measure (i.e. initial plus forced) can be termed 'total identifications'. This total number of identifications measure was used, as it represented a forced-choice paradigm comparable with the analysis used for the C-A calculations in Experiments 6 and 7. There were thus two identification measures, after hypnosis, 1) the 'initial identifications', which were not forced choice, and were comparable to the identification measures used in Experiments 1 to 5, and 2) the 'total identifications', which were forced choice and were comparable to the measures used in Experiments 6 and 7.

The number of 'initial identifications' (i.e. not forced) that subjects made and the average confidence that they expressed in identifications after hypnosis were analyzed with 2 X 2 ANOVAs (verified/unverified X high/low hypnotisability) (see Table, 17.3.2 for means and SDs). No significant differences were found in the number of identifications that subjects made with respect to verified/unverified group $F(1,23)=3.71$ $p>.07$. However, there was a significant effect of hypnotic susceptibility. Subjects of high susceptibility made significantly more identifications than those of low hypnotic susceptibility, $F(1,23)=5.11$ $p<.03$. There was no significant interaction between verified/unverified group and high/low hypnotic susceptibility, $F(1,23)=0.91$ $p>.35$.

No significant differences were found in the average confidence that subjects expressed in their total identifications between verified/unverified group

$F(1,23)=0.31$ $p>.58$. However, there was a significant effect of hypnotic susceptibility. Subjects of high susceptibility expressed significantly greater confidence in their identifications than those of low hypnotic susceptibility, $F(1,23)=9.45$ $p=<.006$. There was no significant interaction between verified/unverified group and high/low hypnotic susceptibility, $F(1,23)=2.44$ $p>.13$.

Table 17.3.2. Verified and unverified group performance before and after hypnosis.

measure	before hypnosis		after hypnosis	
	verified N=13	unverified N=14	verified N=13	unverified N=14
estimated no.	13.92 (2.99)	12.29 (4.78)	15.23 (4.60)	15.07 (4.84)
conf. in estimated no.	4.62 (1.39)	4.93 (1.14)	4.38 (0.96)	5.43 (1.09)
identify in court	10.00 (3.34)	9.29 (4.18)	11.00 (4.00)	12.29 (5.20)
number of identifications			20.46 (3.93)	17.00 (5.86)
average confidence in total identifications			2.98 (0.64)	3.12 (0.86)

Note: Standard deviations in brackets

Table 17.3.3. Performance before and after hypnosis of high and low hypnotically susceptible subjects.

measure	before hypnosis		after hypnosis	
	low N=12	high N=15	low N=12	high N=15
estimated no.	14.00 (3.28)	12.33 (4.22)	17.00 (4.13)	13.67 (4.61)
conf. in estimated no.	4.75 (0.97)	4.80 (1.47)	5.00 (1.41)	4.87 (0.92)
identify in court	9.92 (4.14)	9.40 (3.52)	12.67 (5.09)	10.87 (4.21)
number of identifications			21.25 (3.77)	16.60 (5.42)
average confidence in total identifications			2.73 (0.51)	3.46 (0.83)

Note: Standard deviations in brackets

Pearson's correlations were calculated for the verified and unverified group independently before and after hypnosis, between subjects' estimates of how many photographs they could accurately identify, the confidence shown in the accuracy of those estimates, the number of identifications that subjects said they would identify in court as well as the number of identifications that subjects attempted, their average confidence in their overall identifications (total identifications) and hypnotisability. These correlations are shown in Tables 17.3.4 and 17.3.5.

The majority of significant correlations were significant for both the verified and the unverified group and were as follows: subjects' estimates of how many photographs they could accurately identify before and after hypnosis (r=.74 and r=.55, respectively); subjects' estimates of how many photographs they could accurately identify before hypnosis and the number of photographs that they said that they would identify in court before hypnosis (r=.62 and r=.82, respectively); subjects' estimates of how many photographs they could accurately identify before hypnosis and their average confidence in identifications (r=.64 and r=.62, respectively); the number of identifications that subjects were prepared to make in court before hypnosis and the number of identifications that subjects were prepared to make in court after hypnosis (r=.88 and r=.69, respectively); subjects' estimates of how many photographs they could accurately identify after hypnosis and the number of identifications that subjects were prepared to make in court after hypnosis (r=.86 and r=.71, respectively); subjects' estimates of how many photographs they could accurately identify after hypnosis and their average confidence in identifications (r=.69 and r=.77, respectively).

However, some correlations were significant for the verified group but not the unverified group and vice versa. Those significant for the verified group only were as follows: subjects' estimates of how many photographs they could identify before hypnosis and the number of identifications that they were prepared to make in court after hypnosis (r=.80); the number of identifications that subjects said they would make in court before and after hypnosis (r=.81); the number of identifications that subjects said they would make in court before hypnosis and their average confidence (r=.77); the number of identifications that subjects said that they would make in court after hypnosis and their average confidence ($r=.76$); and between hypnotisability and average confidence ($r=.68$). While for the unverified group the only significant correlations were as follows: subjects' estimates of how many accurate identifications they could make after hypnosis and the number of identifications that they made ($r=.63$); the number of identifications that subjects made and their average confidence ($r=.58$); subjects' confidence in the accuracy of the estimated number of photographs that they could accurately identify after hypnosis and hypnotisability ($r=.61$); and the number of identifications that subjects made and hypnotisability ($r=.64$).

Table 17.3.4. Correlations (*r*) between performance variables for the verified group, ($N = 19$)

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1. estimated no.	-	-.19	.62*	.74**	-.08	.80**	.18	.64*	.38	-.07	.30	.18	.52
2. conf. in 1.		-	.13	.18	.31	.08	-.50	.03	-.17	-.48	.12	.36	-.05
3. identify in court			-	.81***	.23	.88***	-.23	.77**	.16	-.41	.12	.35	.30
4. estimated no.				-	-.12	.86***	-.20	.69**	.06	-.30	.12	.25	.39
5. conf. in 4.					-	.07	-.16	.12	.07	-.26	.18	.19	.08
6. identify in court						-	.06	.76**	.24	-.12	.39	.11	.43
7. no. of identifications made							-	.19	.59	.78**	-.07	-.62*	.52
8. overall confidence								-	.63*	-.26	.45	.32	.68*
9. no. correct identifications (initial)									-	-.05	.69**	.25	.71**
10. no. incorrect ids (initial)										-	-.62**	.96***	.09
11. no. correct (total)											-	.78**	.39
12. accuracy rate (initial)												-	.07
13. hypnotic susceptibility													-

Note:- * $p < .05$, ** $p < .01$, *** $p < .001$ *.Items 1, 2 and 3 are before motivating instructions. Items 4-10 are after motivating instructions.*

Table 17.3.5. Correlations (r) between performance variables for the unverified group, N=20

Measure	1	2	3	4	5	6	7	8	9
1. estimated no.	-	-.46	.82***	.55*	-.48	.48	-.10	.62*	-.04
2. conf. in 1.		-	-.19	-.04	.40	.26	.16	-.07	.04
3. identify in court			-	.35	-.37	.69**	.10	.49	-.10
4. estimated no.				-	-.02	.71**	.63*	.77**	.46
5. conf. in 4.					-	-.09	.39	-.06	.61*
6. identify in court						-	.52	.57	.22
7. no. identifications made							-	.58**	.64*
8. overall confidence								-	.35
9. hypnotic susceptibility									-

Note:-
*$p<.05$, **$p<.01$, ***$p<.001$.
Items 1, 2 and 3 are before motivating instructions. Items 4, 5 and 6 are after motivating instructions.

17.3.2 Actual performance of verified group: initial identifications

Further analysis, considering actual performance, was possible for the verified group. This was further divided into two sub-sections 1) comparisons between subjects' estimates of their accuracy and their number of correct identifications, and 2), confidence-accuracy relationships.

Preliminary analyses included hypnotic depth (hypnotisability) as a variable; however, none of the main effects or interactions involving this variable were significant, with the exception of that relating to correct answers, and this is more accurately assessed in the correlational analysis which is presented shortly. Hence hypnotic depth is excluded from the following analyses.

Comparisons were made between subjects' estimates of their accuracy and the number of correct identifications that they made (based on the 'initial' identifications'). Subjects in the verified group made an average 15.23 (SD= 2.50) initial correct identifications and 5.23 (SD= 3.06) initial incorrect identifications. The initial accuracy rate was .76 (SD= .12).

A one-way repeated measures ANOVA was conducted comparing subjects' estimates of how many accurate identifications they could make both before and after hypnosis and the actual number of initial correct identifications. There was no significant difference between subjects' estimates of how many accurate identifications they could make and the number of initial correct identifications that they made $F(2,24)=1.12$ $p>.34$.

However, when a one-way repeated measures ANOVA with three levels was conducted that compared the number of identifications subjects stated that they would identify in court before and after hypnosis, and the number of correct initial identifications, a significant difference was found $F(2,24)=18.02$ $p<.0001$. Follow-up F tests ($p<.05$) showed that the number of identifications that subjects stated that they would identify in court before and after hypnosis were not significantly different from one-another. However, the number of identifications that subjects stated they would make in court both before and after motivating instructions was less than the actual number of correct identifications that they made (all $ps<.05$).

A 2 X 2 ANOVA with repeated measures on both factors (subjects' estimates of the number of photographs they could accurately identify/the number of identifications subjects stated that they would identify in court X before/after hypnosis) indicated that subjects made significantly higher estimates of the number of photographs that they could accurately identify than the number that they said they could identify in court, $F(1,12)=67.48$ $p<.0001$. There was a significant increase in the number of identifications that subjects estimated that they could accurately achieve and were prepared to testify to in court after hypnosis, $F(1,12)=5.17$ $p<.05$. There was no significant interaction, $F(1,12)=0.09$ $p>.77$. Significant Pearson's correlations were found between the following variables (for the full matrix see Table 17.3.4): the number of correct (initial) identifications and the number of incorrect identifications ($r=.78$); the number of correct (initial) identifications and the number of correct (total) identifications ($r=.69$); average confidence and the number of correct (initial) identifications ($r=.63$); the number of identifications made and accuracy rate ($r=.62$); the number of incorrect (initial) identifications and the number of correct (total) identifications ($r=-.62$); number of incorrect (initial) identifications and accuracy rate (initial) (r=-.96); number of correct identifications (total) and accuracy rate (initial) (r=.78); and the number of correct (initial) identifications and hypnotic susceptibility (r=.68).

17.3.3. Actual performance of verified group: total identifications

A C-A correlation was then calculated for each subject across all 25 identifications. This resulted in 13 within-subjects correlations. These correlations were then averaged. The average correlation was r=.52, SD=.15. To test whether this correlation was significantly different from zero the same procedure was used as for Experiments 6 and 7. The result was significant, $t(12)=2.83$, $p<.01$. Furthermore, each subject's average accuracy was correlated with his/her average confidence rating, this correlation was not significant, r=.33 $p>.27$.

A between subjects C-A correlation was calculated for each identification that subjects made (i.e. C-A correlations were calculated across subjects for the first identification that subjects made, a C-A correlation was calculated across subjects for the second identification that subjects made, etc. and these correlations were averaged). This correlation was r=.32, SD=.23 which was not significantly different from zero $t(23)=1.44$ $p>.10$. In addition, the average accuracy and average confidence score for each identification number was correlated (i.e. the average confidence expressed in the first identification and the average accuracy of the first identification, the average confidence expressed in the second identification and the average accuracy of the second identification etc. were correlated). This correlation was high and very significant, $r=.91$, $p<.0001$.

Subjects' average confidence scores were calculated for correct identifications and incorrect identifications. Subjects expressed significantly greater confidence in correct identifications than in incorrect identifications, $F(1,11)=130.72$ $p<.0001$ (mean confidence in correct identifications= 3.40, SD= 0.79; mean confidence in incorrect identifications= 1.93, SD= 0.70).

Out of a total of 325 identifications, 95 of the identifications given by subjects were rated as being 'absolutely certain' that their identifications were correct. Of these 95 identifications 92 were correct, an accuracy rate of 96.8%.

The influence of hypnotisability was also investigated with respect to confidence and accuracy. Significant correlations were found between hypnotisability and confidence in incorrect identifications ($r=.77$ $p<.002$) but not between hypnotisability and confidence in correct identifications ($r=.31$ $p>.31$) or hypnotisability and within-subjects C-A correlations ($r=-.35$ $p>.24$).

17.4 Discussion

17.4.1 Verified/unverified group comparisons

The initial hypotheses with regard to verifiability were: 1) subjects whose answers could not be verified would estimate that they could accurately identify significantly more photographs than subjects whose answers could be verified; 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimate of the number of photographs that they could accurately identify than subjects whose answers could be verified; 3) subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court than subjects whose answers could be verified; and, 4) that subjects whose answers could not be verified would express greater confidence in the identifications that they made than subjects whose answers could be verified.

None of these hypotheses was supported, although there were trends in the hypothesised direction with respect to the confidence that subjects expressed in the accuracy of the number of photographs that they estimated that they could identify, and for the number of identifications that subjects made.

The hypnotic instructions did, however, appear to have some effects. Subjects showed increases in the number of identifications that they estimated that they could accurately identify and the number of identifications that they would make in court after hypnosis. However, in this respect they simply mirrored the effects shown in Experiments 2 and 4.

Hypnotic susceptibility appeared to influence subjects' performance. Subjects of high susceptibility made significantly more identifications than those of low hypnotic susceptibility, and expressed significantly greater average confidence in their identifications. This would appear to fit in well with other hypnosis findings (see Wagstaff, 1993, 1995). In addition, hypnotic susceptibility was significantly correlated with average confidence in identifications in both the verified and unverified groups, and subjects' confidence in the accuracy of their estimate of the number of photographs that they identified after hypnosis, and the number of identifications that they made in the unverified group. In accordance with socio-cognitive accounts of hypnosis, perhaps these findings occurred because subjects that were of high hypnotic susceptibility were those that most wished to 'help out' the experimenter irrespective of whether the task in question was saying that they were hypnotised, paying attention to stimuli, making a large number of identifications, or expressing high confidence. Indeed, it is plausible that what hypnotic susceptibility is measuring here is subjects' willingness to aid the experimenter/hypnotist (see Wagstaff, 1981). If such an hypothesis is correct, one might expect subjects of high hypnotic susceptibility to make more identifications and express high confidence regardless of whether hypnosis is used as an interview technique. Alternatively, perhaps this willingness to help out only manifests itself in the hypnosis situation.

Correlations between performance measures for the verified group and the unverified group were broadly similar. Generally, subjects who made high estimates of the number of photographs that they could identify appeared to do so consistently, whether the measure was their estimates of how many photographs they could accurately identify, the number they would identify in court and regardless of whether this was before or after hypnosis. These subjects also tended to make more identifications and express greater confidence in individual identifications.

17.4.2 Actual performance of verified group: 'initial identifications'

A subsidiary hypothesis was that there would be a positive relationship between subjects' estimates of how many photographs they thought they could accurately identify and the number of correct identifications that they made. The appropriate correlations were not significant, nevertheless, comparisons between subjects' estimates of how many accurate identifications they could make before and after hypnosis and the actual number of correct identifications that subjects made after hypnosis did not differ. Thus, only when subjects' estimates were measured as a group was there some support for this hypothesis. This is consistent with Experiments 1, 2, 3 and 4.

When the number of identifications that subjects stated that they would identify in court before and after hypnosis were compared with the number of correct identifications that they made, all three figures were different from one another. Although subjects increased the number of photographs that they stated that they would identify in court after hypnosis, this estimate was still significantly less than the number of photographs that they could identify. Thus, although the number of photographs that subjects stated that they would identify in court was increased by hypnosis, it was still a conservative estimate, compared to the actual number of identifications that subjects could make. These findings are again consistent with those of Experiments 2 and 4.

Comparisons between subjects' estimates of the number of photographs they could accurately identify and the number of identifications subjects stated that they would identify in court before and after hypnosis indicated that subjects made significantly higher estimates of the number of photographs that they could accurately identify than the number that they said they could identify in court. Once again this supports previous findings which suggest that the number of identifications that subjects say they would identify in court is a conservative measure.

The number of identifications made correlated with the number of incorrect identifications but not the number of correct identifications and thus negatively correlated with initial accuracy rate. This latter effect was not significant in the previous studies and would seem to suggest that subjects who made a greater number of identifications were lowering their criterion for report. Nevertheless, the number of initial incorrect identifications made was negatively correlated with the number of total correct identifications and therefore also negatively correlated with total accuracy rate. Also the proportion of correct to incorrect responses was, once again, 76/24.

17.4.3. Actual performance of verified group: total identifications

Investigation of the C-A relationships using the total identifications data showed a number of significant positive effects. A significant, within subjects C-A relationship was found. In addition, average accuracy and average confidence score for each identification number was significantly correlated. In addition, the between-subjects C-A correlation was positive, as was the correlation between each subjects' average confidence and each subjects' average accuracy. However, neither of these correlations were significant. Further support for a strong within-subjects C-A relationship was indicated by the finding that subjects expressed significantly greater confidence in correct identifications than incorrect identifications. Also, once again, when subjects indicated that they were 'absolutely certain' in the accuracy of a response they were very likely to be accurate. These results seem to broadly support the previous findings of Experiments 6 and 7. When subjects are confident they tend to be accurate.

Interestingly, hypnotisability was significantly correlated with confidence in incorrect identifications but not with confidence in correct identifications. This would appear to fit with previous findings that indicate that hypnosis may spuriously inflate confidence. Nevertheless, hypnotisability was positively significantly related to the number of initial correct identifications, and positively, though just not significantly (.39) to the number of correct total identifications. It was not related, however, to either measure of incorrect identifications. Thus notwithstanding the results for confidence, if one assumes the logic that 'only the high susceptibles were hypnotised', 'hypnosis' did appear to improve performance to some extent. There are, however, alternative ways of interpreting this finding. Perhaps rather than improving the performance of the high susceptibles, hypnosis worsens the performance of the low susceptibles; i.e. low susceptibles may feel negatively towards a situation defined to them as one of hypnosis, and not respond as well as they might in a more neutral situation (Wagstaff, 1981).

In sum, this experiment indicates that, in a situation defined as one of hypnosis, there was still a positive relationship between subjects' confidence and their accuracy regardless of whether this was a comparison of their various estimates of their ability to make identifications or their direct judgements of how confident they were in the accuracy of each identification that they provided. However, hypnotic susceptibility appeared to mediate this relationship, increasing confidence in incorrect answers whilst also increasing the number of correct responses. It was not clear from the paradigm used, however, whether this was because of characteristics associated with hypnotic susceptibility, the fact that the situation was defined as one of 'hypnosis' or a combination of both. Also, as in the previous experiments no attempt was made to determine whether the memory facilitation instructions improved actual memory. Although previously it was argued that providing verified subjects with an opportunity to make identifications before the memory facilitation/motivating instructions might introduce confounding effects, in view of the relative paucity of any differences between the verified and unverified conditions, and the importance of making a before/after estimate of actual accuracy, it was decided to throw caution to the wind and attempt such an estimate in the next experiment.

CHAPTER 18: EXPERIMENT 9

AN EXPERIMENTAL INVESTIGATION OF THE EFFECTS OF VERIFICATION, INTERVIEW INSTRUCTIONS AND CONFIDENCE ON A FACE RECOGNITION TASK AND A RECALL TASK.

18.1 Introduction

Experiment 8 showed little effect of hypnosis on the performance of a verified group compared to an unverified group. This may have implications for the cognitive interview per se (i.e. in the absence of hypnosis), as many of the criticisms that have been applied to hypnosis have also been applied to techniques like the cognitive interview. For instance, it could be argued that during such procedures subjects may feel that they are under pressure to recall more information, because it is clear why the procedure is being given, i.e. to enhance recall. Also the use of imagery may encourage subjects to fantasize and make things up (see for example, Perry and Nogrady, 1985). This may lead them to produce exaggerated memory reports, particularly when answers cannot be verified. However, although the cognitive interview instructions were used in conjunction with hypnosis in Experiment 8 and no differences between verified and unverified groups were apparent, it is not inevitable that the cognitive interview without hypnosis would fare better or no worse than with hypnosis in these respects. For example, possibly the pressure to recall more information from the cognitive interview instructions might be offset by the instructions to relax in the hypnosis procedure, or the hypnosis procedure might encourage more truthful reporting.

The influence of context-reinstatement itself on identification accuracy has been controversial (see for example, McSpadden, Schooler, & Loftus, 1988). While a number of studies have shown positive effects (e.g. Cutler, Penrod & Martens, 1987; Krafka & Penrod, 1985: Malpass & Devine, 1981; Timm, 1981; Wagstaff, 1982) others have not (e.g. Lindsay & Wallbridge, 1983; McSpadden, Schooler, & Loftus, 1988). Cutler, Penrod and Martens (1987) have suggested that these discrepant findings may be depend on the effectiveness of additional contextual cues given to subjects to enhance recall and recognition compared to the cues that are already available to the subject. With respect to identification tasks in particular, they suggest that the target itself will act as a strong contextual cue, and thus, provision of other contextual cues will have little effect. However, they do suggest that exceptions may occur, for example if there has been a long delay since the target was shown, or disguises were used.

Cutler et al. have also suggested that discrepancies between different context-reinstatement studies may be explained by the different contextual cues have been used in different experiments (although they also note that even when similar contexts have been used some positive and some null effects have been reported). McSpadden et al. suggest that their own failure to achieve robust context effects may be due to the presentation of context-reinstatement instructions by tape and because they did not use multiple-context reinstatement techniques (as used by, for example, Cutler et al., 1987; Geiselman et al., 1984; Malpass & Devine, 1981). Given this controversy, a comparison of different interview conditions that involve multiple-context reinstatement and tape-recorded interview instructions, on performance of a face-recognition task, would seem appropriate.

Surprisingly, little research has actually been conducted specifically comparing hypnosis with the cognitive interview. Although some researchers have compared hypnosis with guided imagery, there is only one published account of a comparison of hypnosis and cognitive interview procedures; that is a study by Geiselman et al. (1985). This found no differences between the two procedures in terms of the recall of correct and incorrect information, though both were better than a standard interview. However, no confidence-accuracy measures were taken in this study, and there was little uniformity in, or detail about, what constituted a 'hypnotic interview'. It is possible, for instance, that in many respects the procedures might have overlapped (see Wagstaff, 1982), and underplayed the 'hypnotic' element. Consequently, if the cognitive interview is to be accepted as a viable alternative to hypnosis, a more definitive demonstration of the possible advantages of the cognitive interview over hypnosis, in terms of increases in accurate information without increases in errors, or enhanced C-A relationships would be useful.

In addition, hypnotic susceptibility appeared to influence performance in Experiment 8. However, because of the absence of an appropriate control group it is difficult to determine if the effects were due to subjects' hypnotic susceptibility *per se*, the fact that the situation was defined as 'hypnosis', or a combination of the two.

Given these considerations a final experiment was conducted that attempted to assess the interactive effects of verifiability, with hypnosis, the cognitive interview, and a control condition, on measures that included confidence and accuracy, and measures of accuracy before and after the memory facilitation procedures. As C-A relationships seemed to be affected by item difficulty, this was also included as a variable.

18.1.1 Hypotheses

The following main predictions were made.

1) Subjects whose answers could not be verified would estimate that they could identify significantly more photographs than subjects whose answers could be verified.

2) Subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimates of how many photographs that they could accurately identify, and the accuracy of their actual identifications, than subjects whose answers could be verified.

3) Subjects whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court than subjects whose answers could be verified.

4) Within the verified group there would be a positive relationship between subjects' estimates of how many photographs that they thought they could accurately identify and the number of correct identifications they made.

5) There would be a large, positive relationship between eyewitness confidence and accuracy if item difficulty is varied by including easy, medium and hard questions.

No specific predictions were made regarding the interview conditions and, hypnotic susceptibility, as any number of different hypotheses could be forwarded; rather the aim was to generally assess the relative impact of these variables on the measures.

18.2 Method

18.2.1 Subjects

A total of 53 subjects, 34 female and 19 male of various backgrounds participated in the experiment. The mean age was 22.60 years (range 18-52, SD= 6.19). Equal numbers of subjects were randomly assigned to one of three *conditions* ('cognitive interview', 'hypnosis' or 'control'). Each of the three conditions was further divided into two *groups* of equal size; 'verified' or 'unverified'.

18.2.2 Materials and Procedure

The procedure used in this experiment was in many respects a combination of Experiments 7 and 8. Subjects were tested in groups, ranging in size from one to five persons. Subjects were randomly assigned to either verified and unverified groups, and one of three conditions; hypnosis, cognitive interview and control. Again, these groups were presented with a 'pack' of fifty photographs of faces which they were required to shuffle. Subjects were then instructed to choose 25 photographs at random from the pack and were asked to look at these for 30 seconds.

The most important part of this procedure was to try to ensure that subjects in the 'unverified' group were aware that their answers could not be verified. This was achieved in the same way as in previous experiments. The unverified group were told to place the photographs which they had chosen back with the photographs from the original 'pack' and to shuffle all the cards together, then they placed their shuffled 'packs' in identical plain envelopes which were placed by the subjects themselves in a box passed around the room.

In contrast, after the timed period viewing the photographs, the 'verified' group placed the photographs which they had chosen separately into an envelope and were told explicitly by the experimenter that this was so that their answers to any subsequent questions could be verified later.

Subjects were then shown a five and a half minute black and white video film that concerned the implied murder of a male by a female that was the used as the stimulus in Experiment 7.

After the film subjects were tested on their recognition of the photographs that they had previously chosen using an identification sheet which they were informed contained all 50 photographs from which they had chosen 25 and the Questionnaire (Questionnaire 1b). Next subjects were required to fill out an Answer Sheet (Answer sheet 1b). (N.B. This is the first time the Answer Sheet was given **before** the memory facilitation/motivating instructions). The Answer Sheet was identical to that used in Experiment 8 (Answer Sheet 1b). When subjects had stopped completing the answer sheet, some subjects had not made 25 identifications on the answer sheet. Further instructions were given to those subjects to draw a line 'under their last response and to fill in the remaining spaces with photographs which they thought were most likely to have been present'

Once these tasks had been completed subjects in the three conditions received the following procedures:

a) hypnotic condition. Subjects were given the following information:

> One method that the Police use that may improve memory is hypnosis. I will now use a hypnotic procedure.

Subjects were then played a taped hypnotic induction procedure. As in the last experiment this was a slightly modified version of the induction procedure provided by Barber (1969) for use with the Barber Suggestibility Scale. Subjects were required to report their state of hypnotic depth according to the modified Long Stanford Scale (Tart, 1970), also was used in Experiment 8.

After this hypnotic induction subjects were given four instructions for memory facilitation. These were based on the four mnemonics used by Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian & Prosk (1984) in the cognitive interview. These instructions were:

1) Reinstate context. 'What I would like you to do now is to think carefully about the photographs and the film that you saw. Think about what you felt and what you thought when you looked at the photographs and the film. Think about how you were feeling at the time, and of your reactions to the photographs and to the film. Try to reinstate the context in your mind, of the physical environment in which you saw the photographs and the film, such as the location of objects and people in the room.'

2) Report everything. 'It is known that some witnesses hold back information- because they are not sure about what they can remember or don't know if the information is relevant. However, you must try to report everything that you can'.

3) Recall the events in different orders. 'Most people remember details of an event in a certain order, from beginning to end. However, you should also try to recall the photographs and the film that you saw in a different order- perhaps starting at the end of the period that you viewed the photographs or the end of the film and working backwards. Alternatively, you can start at a photograph or part of the film that you can remember particularly well and work either forwards or backwards from there'.

4) Change perspectives. 'People who witness events sometimes try to remember events from somebody else's perspective, so try to remember what you would have seen if you were a different person viewing the photographs or what you would have felt and seen if you were one of the characters in the film'.

As in the previous experiment subjects were then instructed to open their eyes, whilst remaining hypnotized and to answer the questionnaire devised to test their recall of the film (Questionnaire 3). The format of the questionnaire was identical to that of Experiment 7. Henceforth, to avoid confusion between the two questionnaires, Questionnaire 1b will be referred to as the 'identification questionnaire' because it tested subjects' ability to identify the photographs that they had seen. Questionnaire 3, because it tested subjects' ability to recall the film will be called the 'recall questionnaire'.

Subjects were then required to repeat the Identification Questionnaire and Answer Sheet concerning the photographs that they had chose originally. They were instructed that after they had completed the questionnaire and the answer sheet they were to close their eyes, whilst remaining hypnotised and await further instructions. (It can be noted that only the identification questionnaire was administered both before and after the memory facilitation instructions; this was because of time constraints. The whole procedure would have exceeded an hour if the recall questionnaire had been administered twice.)

When all subjects had completed this task and had closed their eyes, some subjects had not made 25 identifications on the answer sheet. Further instructions were given to those subjects to open their eyes, to draw a line under their last identification and carry out the following instruction:

> 'if you are not at all sure which of the remaining photographs were the ones that you chose, fill in the remaining spaces with photographs which you think were most likely to have been present.'

Subjects who had made 25 identifications were instructed to:

> 'remain relaxed with your eyes closed.'

On final completion of the answer sheet, subjects were then 'woken up' by counting from five to one, as before. Subjects were then debriefed and thanked for their participation.

b) Cognitive Interview condition. Subjects were first given a reading filler activity designed to take the same time as the hypnotic induction procedure (this was the same as in Experiment 1; i.e. an article about hip replacements and car seat-belts). They were then given the following information.

> One method that the Police use that may improve memory is a cognitive interview. I will now use a cognitive interview.

Subjects were then played an tape recording of exactly the same four memory mnemonics which were used with the hypnotic group. As explained previously, this was so that 'hypnosis' per se would not be confounded with differences in memory facilitation instructions between the two groups.

After these instructions subjects were required to complete the recall questionnaire. Then subjects were required to repeat the identification questionnaire and answer sheet. After they had stopped filling in the answer sheet, again some subjects had not made 25 identifications on the answer sheet. Subjects were told to draw a line under their last identification and carry out the following instruction.

> if you are not at all sure which of the remaining photographs were the ones that you chose, fill in the remaining spaces with photographs which you think were most likely to have been present.

On completion of the answer sheet, subjects were then given the same hypnotic induction procedure as in the hypnotic condition. They were also required to rate their depth of hypnosis before being 'woken up'. Finally they were debriefed and thanked for their participation.

c) Control condition. The procedure for the control condition was identical to that of the cognitive interview condition but without the mnemonic instructions. The filler activity was increased by 3 minutes so that there was a similar delay between stimuli presentation and memory testing as in the other groups. After the filler activity subjects were instructed as follows.

> One method that the Police use that may improve memory is asking witnesses to try hard. Please try hard to remember the film and the photographs.

After this they were tested inexactly the same way as the cognitive interview condition, including an assessment of hypnotic susceptibility.

18.3 Results

The results section is divided into three sections 1) subjects' performance on the identification questionnaire task, 2) subjects' performance on the recall questionnaire task, and 3) comparisons between the face recognition task and the recall questionnaire task.

18.3.1 Identification questionnaire

The results for the face recognition task are further divided into two parts 1) verified/unverified comparisons, and 2) actual performance, in terms of correct and incorrect identifications and C-A relationships, of the verified group.

18.3.1.1 Verified/unverified group comparisons

Depth scores are shown in Table 18.3.1.1. A one-way ANOVA (control/cognitive interview/hypnosis) on these hypnotic depth scores revealed no significant effect for interview condition, $F(2,49)=2.48$ $p>.09$; however, it can be noted that the scores, tend to be slightly higher in the hypnosis condition, perhaps because they received the induction earlier in the procedure. The range of these scores was from 0 to 6. As in the last experiment, subjects were divided into high and low susceptibility using an LSS score of 3 as the cut off.

Table 18.3.1.1. Hypnotisability of subjects with respect to interview condition.

	interview condition		
control	cognitive interview	hypnosis	combined
2.18 (1.47)	1.82 (1.29)	3.06 (2.15)	2.37 (1.74)

Note: Standard deviations in brackets.

The performance of verified/unverified groups was compared in terms of 1) subjects' estimates of the number of photographs that they could accurately identify, 2) the confidence expressed in the accuracy of that estimate, 3) the number of identifications that they would be prepared to testify in court to have seen before, 4) the number of identifications that subjects made, and 5) the average confidence that they expressed in identifications that they had made. The means and SDs for these analyses are displayed for the control condition, the cognitive interview condition, the hypnosis condition and combined in Tables 18.3.1.2, 18.3.1.3, 18.3.1.4, and 18.3.1.5, respectively.

The above variables were analyzed with 3 X 2 X 2 ANOVAs with repeated measures on the third factor (control/cognitive interview/hypnosis X verified/unverified X before/after interview instructions). Hypnotic susceptibility, was originally included as a factor in the ANOVAs, however, it was dropped as it showed no significant effects. The results were as follows.

No significant differences were found in the number of photographs that subjects estimated that they could identify with respect to interview condition, $F(2,47)=0.09$ $p>.91$, verified/unverified group $F(1,47)=0.32$ $p>.57$, or before/after interview instructions $F(1,47)=0.40$ $p>.67$. The interactions between interview condition and verified/unverified group, between interview condition and before/after interview instructions, between verified/unverified groups and before/after interview instructions, and between interview condition, verified/unverified group and before/after interview instructions were all not significant, $F(2,47)=0.27$ $p>.76$, $F(2,47)=0.40$ $p>.67$, $F(1,47)=1.16$ $p>.29$ and $F(2,47)=1.08$ $p>.35$ respectively.

No significant differences were found in the confidence expressed in this estimate with respect to interview condition $F(2,47)=0.73$ $p>.49$, verified/unverified group $F(1,47)=3.71$ $p>.06$ or before/after interview instructions $F(1,47)=0.01$ $p>.92$. The interactions between interview condition and verified/unverified group, between interview condition and before/after interview instructions, between verified/unverified groups and before/after interview instructions, and between interview condition, verified/unverified group and before/after interview instructions were all not significant, $F(2,47)=2.12$ $p>.23$, $F(2,47)=0.32$ $p>.73$, $F(1,47)=0.01$ $p>.92$ and $F(2,47)=0.81$ $p>.45$ respectively.

No significant differences were found in the number of identifications that subjects stated that they would identify in court with respect to interview condition $F(2,47)=0.39$ $p>.68$, verified/unverified group $F(1,47)=0.29$ $p>.59$ or before/after interview instructions $F(1,47)=1.56$ $p>.22$. The interactions between interview condition and verified/unverified group, between interview condition and before/after interview instructions, between verified/unverified groups and before/after interview instructions, and between interview condition, verified/unverified group and before/after interview instructions were all not significant, $F(2,47)=0.93$ $p>.40$, $F(2,47)=0.82$ $p>.45$, $F(1,47)=1.59$ $p>.23$ and $F(2,47)=0.59$ $p>.56$ respectively. No significant differences were found in the number of initial identifications that subjects made with respect to interview condition $F(2,47)=0.71$ $p>.50$ or group $F(1,47)=0.05$ $p>.82$. However, there was a significant increase in the number of identifications that were made before/after interview instructions had been given $F(1,47)=13.40$ $p<.0006$. The interactions between interview condition and verified/unverified group, between interview condition and before/after interview instructions, between verified/unverified groups and before/after interview instructions, and between interview condition, verified/unverified group and before/after interview instructions were all not significant, $F(2,47)=0.40$ $p>.68$, $F(2,47)=0.96$ $p>.39$, $F(1,47)=0.00$ $p>.95$ and $F(2,47)=1.07$ $p>.35$ respectively. With respect to total identifications no significant differences were found in the average confidence expressed in identifications that subjects made with respect to interview condition, $F(2,47)=0.38$ $p>.68$, or verified/unverified group $F(1,47)=0.02$ $p>.90$. However, there was a significant increase in the overall confidence expressed in identifications before/after the interview instructions, $F(1,47)=41.82$ $p<.0001$. A significant interview condition by before/after interview instructions interaction was also found, $F(2,44)=7.63$ $p<.002$. One-way between subjects ANOVAs (control/cognitive interview/hypnosis) indicated that this was not due to differences between conditions before or after interview procedures were applied, $F(2,44)=0.00$, $p>.99$ and $F(2,44)=1.47$, $p>.24$ respectively. Follow up F tests on the within subjects comparisons showed significant increases in average confidence before/after interview instructions in the control condition and the hypnosis condition but not for the cognitive interview condition. As an alternative way of conceptualising the data, difference scores were calculated by subtracting subjects' average confidence scores before interview instructions from the corresponding scores after interview instructions. A one-way ANOVA (control/cognitive interview/hypnosis) on the difference scores between each subject's average confidence before/after interview instructions revealed a significant effect of condition, $F(2,47)=8.12$ $p<.0009$. Follow-up F tests indicated that the hypnotic group (**M**=13.00, **SD**=8.12) had a significantly greater increase than both the control (**M**=3.89, **SD**=6.07) and the cognitive interview group (**M**=3.76, **SD**=7.94). The cognitive interview group was not significantly different from the control group. The remaining interactions between interview condition and verified/unverified group, between verified/unverified groups and before/after interview instructions, and between interview condition, verified/unverified group and before/after interview instructions were not significant, $F(2,47)=0.12$ $p>.89$, $F(2,47)=0.93$ $p>.34$, and $F(1,47)=0.80$ $p>.46$, respectively.

Table 18.3.1.2. Verified and unverified group performance before and after control instructions.

measure	before instructions		after instructions	
	verified N=9	unverified N=9	verified N=9	unverified N=9
estimated no.	11.89 (6.29)	11.00 (1.41)	9.94 (6.98)	11.44 (3.05)
conf. in estimated no.	4.33 (0.71)	4.67 (0.87)	4.22 (1.09)	5.22 (0.44)
identify in court	7.89 (4.37)	5.67 (2.38)	7.89 (5.01)	5.33 (2.83)
number of identifications	19.00 (7.07)	16.56 (5.13)	19.89 (7.99)	19.44 (6.69)
average confidence in total identifications	2.34 (0.65)	2.41 (0.33)	2.47 (0.81	2.57 (0.37)

Note: Standard deviations in brackets.

Table 18.3.1.3. Verified and unverified group performance before and after cognitive interview procedure

measure	before instructions		after instructions	
	verified N=8	unverified N=9	verified N=8	unverified N=9
estimated no.	13.00 (5.86)	10.00 (2.56)	11.38 (7.13)	10.33 (4.97)
conf. in estimated no.	4.25 (1.49)	5.33 (1.12)	4.25 (1.58)	5.22 (0.97)
identify in court	6.50 (5.18)	7.89 (2.03)	7.25 (4.10)	8.33 (3.16)
number of identifications	19.00 (7.07)	16.56 (5.13)	19.89 (7.99)	19.44 (6.69)
average confidence in total identifications	2.33 (0.75)	2.40 (0.46)	2.48 (0.75)	2.56 (0.36)

Note: Standard deviations in brackets

Table 18.3.1.4. Verified and unverified group performance before and after hypnosis procedure.

measure	before instructions		after instructions	
	verified N=9	unverified N=9	verified N=9	unverified N=9
estimated no.	11.56 (4.61)	11.56 (4.95)	12.33 (4.82)	11.33 (5.43)
conf. in estimated no.	4.44 (1.42)	4.44 (1.13)	4.56 (1.59)	4.11 (0.78)
identify in court	5.75 (5.99)	6.06 (3.36)	7.50 (5.71)	5.94 (3.08)
number of identifications	13.89 (6.13)	14.33 (4.87)	20.33 (7.07)	17.22 (6.51)
average confidence in total identifications	2.35 (0.85)	2.36 (0.48)	2.98 (0.88)	2.75 (0.70)

Note: Standard deviations in brackets

Table 18.3.1.5. Verified and unverified group performance before and after interview instructions (control, cognitive interview and hypnosis conditions combined).

measure	before instructions		after instructions	
	verified N=26	unverified N=27	verified N=26	unverified N=27
estimated no.	12.11 (5.43)	10.85 (3.25)	9.94 (6.98)	11.03 (4.45)
conf. in estimated no.	4.34 (1.20)	4.81 (1.08)	4.35 (1.38)	4.85 (0.91)
identify in court	6.76 (5.05)	6.56 (2.71)	7.56 (4.78)	6.56 (3.19)
number of identifications	16.08 (6.58)	15.74 (5.58)	19.26 (7.42)	18.78 (6.47)
average confidence in total identifications	2.33 (0.72)	2.39 (0.40)	2.65 (0.80)	2.61 (0.46)

Note: Standard deviations in brackets

Table 18.3.1.6. Control, cognitive interview and hypnosis performance before and after procedure.

measure	control		cognitive interview		hypnosis		combined	
	before N=18	after N=18	before N=17	after N=17	before N=18	after N=18	before N=53	after N=53
estimated no.	11.44 (4.45)	10.69 (5.28)	11.41 (4.54)	10.82 (5.91)	11.56 (4.64)	11.83 (5.01)	11.47 (4.46)	11.12 (5.32)
conf. in estimated no.	4.50 (0.79)	4.72 (0.96)	4.82 (1.38)	4.76 (1.35)	4.44 (1.25)	4.33 (1.24)	4.58 (1.15)	4.60 (1.18)
identify in court	6.78 (3.61)	6.61 (4.16)	7.23 (3.78)	7.82 (3.56)	5.91 (4.70)	6.72 (4.50)	6.66 (3.99)	7.05 (4.04)
number of identifications attempted	17.78 (6.12)	19.67 (7.15)	15.82 (6.34)	18.59 (7.08)	14.11 (5.38)	18.78 (6.79)	15.91 (6.03)	19.02 (6.89)
average confidence in total identifications	2.36 (0.59)	2.52 (0.61)	2.36 (0.59)	2.52 (0.56)	2.32 (0.68)	2.87 (0.78)	2.36 (0.57)	2.63 (0.65)

Note: Standard deviations in brackets

Pearson's correlations were calculated for the verified and unverified groups independently, for before and after interview instructions, between subjects' estimates of how many photographs they could accurately identify, the confidence shown in the accuracy of those estimates, the number of identifications that subjects said they would identify in court, the number of identifications that subjects made and the average confidence that they expressed in identifications. Correlations for the control condition are displayed in Table 18.3.1.7 for the verified group and Table 18.3.1.8 for the unverified group. Correlations for the cognitive interview condition are displayed in Table 18.3.1.9 for the verified group and Table 18.3.1.10 for the unverified group. Correlations for the hypnosis condition are displayed in Table 18.3.1.11 for the verified group and Table 18.3.1.12 for the unverified group. Tables 18.3.1.13 and 18.3.1.14 show correlations for interview conditions combined for the verified and unverified groups respectively.

Because there were so many correlations (>400) only those of particular note are described here. For the remainder the reader is referred to the appropriate Tables.

Broadly speaking, similar results were found as for previous experiments. Subjects' estimates of their performance seemed to be related to one another for both the verified and unverified group, irrespective of interview instructions. Thus, subjects estimates of the number of accurate identifications that they had made, the number of identifications that they would make in court, the number of identifications made and average confidence before and after interview instructions all tended to be related.

With respect to hypnotisability, overall, there appeared to be a discrepancy between verified and unverified groups. For the verified group subjects' estimate of how may accurate identifications the could make before and after interview instructions, the number of identifications that they would make in court after interview instructions, the number of identifications made before and after interview instructions and the total average confidence in identifications before and after interview instructions were all significantly correlated with hypnotisability. These effects did not appear to be greatly influenced by interview instructions. None of these was significant for unverified group.

Table 18.3.1.7. Correlations (r) between performance variables for the verified group in the control condition.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. estimated no.	-	.21	.89**	.46	.54	.49	.36	.33	-.43	.86**	-.52	.88**	.59	.47	.54	.54	.50	-.54	.70*
2. conf. in 1.		-	.38	-.17	.07	.26	-.34	.23	.39	.22	.38	.29	-.10	.15	.28	-.26	.20	.34	.21
3. identify in court			-	.42	.69*	.75*	.20	.50	-.25	.93*	-.15	.96***	.57	.68*	.73*	.43	.66	-.39	.82**
4.no. identifications made				-	.61	.70*	.95*	.59	-.82**	.41	-.06	.32	.89*	.61	.56	.92***	.48	-.87**	.51
5. average confidence					-	.79*	.41	.37	-.38	.74*	.01	.74*	.82*	.94***	.89**	.68*	.73*	-.56	.72*
6. no. correct identifications (initial)						-	.43	.70*	-.33	.63	.32	.62	.75*	.84**	.83**	.61	.68*	-.49	.83**
7. no. incorrect identifications (initial)							-	.43	-.88**	.24	-.23	.12	.79*	.39	.34	.89**	.30	-.88**	.27
8. no. correct identifications (total)								-	-.57	.32	.41	.32	.48	.59	.62	.36	.77*	-.51	.44
9. accuracy rate (initial)									-	.26	.30	-.20	-.72*	-.43	-.41	-.77*	-.54	.95***	-.24
10. estimated no.										-	-.33	.97***	.59	.70*	.68*	.47	.59	-.39	.67*
11. conf. in 10.											-	-.29	-.20	.16	.12	-.31	.14	.35	-.05
12. identify in court												-	.55	.69*	.72*	.40	.64	-.35	.74*
13. no. identifications made													-	.80**	.80**	.96***	.65	-.89**	.72*
14. average confidence														-	.97***	.62	.87**	-.57	.68
15. no. correct identifications (initial)															-	.61	.91***	-.58	.78*
16. no. incorrect identifications (initial)																-	.45	-.91***	.61
17. no. correct identifications (total)																	-	-.59	.58
18. accuracy rate (initial)																		-	-.48
19. hypnotic susceptibility																			-

Note:-$^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$. *Items 1 - 9 are before control instructions. Items 10 - 19 are after control instructions.*

Table 18.3.1.8. Correlations (r) between performance variables for the unverified group in the control condition.

Measure	1	2	3	4	5	6	7	8	9	10	11
1. estimated no.	-	.20	.48	.17	.52	.29	.00	.63	.46	$.70^{*}$	-.44
2. conf. in 1.		-	.24	.34	.37	.39	.54	.05	-.10	.27	.37
3. *identify in court*			-	-.02	.64	.18	.32	$.87^{***}$	.16	.54	.26
4.no. identifications made				-	.57	-.56	-.17	.10	$.88^{**}$	.50	-.45
5. average confidence					-	-.07	.36	.60	$.68^{*}$	$.86^{**}$	.02
6. estimated no.						-	-.08	.63	-.40	.22	-.37
7. conf. in 6.							-	.13	-.25	.23	$.71^{*}$
8. identify in court								-	.18	$.76^{*}$	-.07
9. no. identifications made									-	.56	-.63
10. average confidence										-	-.28
11. hypnotic susceptibility											-

Note:-
$^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$.
Items 1 - 5 are before control instructions. Items 6 - 10 are after control instructions.

Table 18.3.1.9. Correlations (*r*) between performance variables for the verified group in the cognitive interview condition.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. estimated no.	-	-.30	.77*	.88**	.98***	.89**	.62	.83*	.02	.93***	.19	.98***	.47	.98***	.58	.27	.52	-.02	.47
2. conf. in 1.		-	.07	-.12	-.25	-.34	-.45	-.46	-.84**	-.43	.52	-.29	-.64	-.40	-.64	-.53	-.58	.22	-.60
3. identify in court			-	.65	.81*	.55	.72*	.55	-.32	.82*	.10	.79*	.13	.75*	.06	.19	-.02	-.25	-.12
4.no. identifications made				-	.90**	.97***	.80*	.81*	-.10	.79*	.11	.78*	.53	.88**	.62	.35	.44	-.12	.56
5. average confidence					-	.89**	.68	.84**	-.02	.94***	.08	.94***	.42	.97***	.50	.27	.39	-.05	.46
6. no. correct identifications (initial)						-	.62	.86**	.14	.80*	-.01	.78*	.68	.90***	.77	.46	.58	-.15	.72*
7. no. incorrect identifications (initial)							-	.46	-.66	.53	.38	.56	.03	.58	.07	-.02	-.03	-.03	.04
8. no. correct identifications (total)								-	.18	.87**	-.22	.79**	.54	.90**	.64	.34	.65	-.03	.67
9. accuracy rate (initial)									-	.07	-.62	-.03	.67	.09	.62	.61	.46	-.30	.53
10. estimated no.										-	-.05	.94***	.45	.97***	.48	.34	.44	-.15	.38
11. conf. in 10.											-	.25	-.42	.03	-.20	.59	.02	.52	-.25
12. identify in court												-	.34	.96***	.46	.15	.50	.06	.35
13. no. identifications made													-	.54	.93**	.91**	.64	-.62	.62
14. average confidence														-	.62	.36	.56	-.11	.50
15. no. correct identifications (initial)															-	.69	.83*	-.29	.80*
16. no. incorrect identifications (initial)																-	.32	-.87**	.32
17. no. correct identifications (total)																	-	.09	.69
18. accuracy rate (initial)																		-	.15
19. hypnotic susceptibility																			-

Note:- *p<.05, ** p<.01, *** p<.001. *Items 1 - 9 are before cognitive interview instructions. Items 10 - 19 are after cognitive interview instructions.*

Table 18.3.1.10. Correlations (r) between performance variables for the unverified group in the cognitive interview condition.

Measure	1	2	3	4	5	6	7	8	9	10	11
1. estimated no.	-	.00	.77*	.53	.58	.80**	.05	.88**	-.16	.06	-.29
2. conf. in 1.		-	.24	-.39	.06	-.11	-.08	-.08	-.39	.09	-.31
3. identify in court			-	.28	.28	.57	.14	.71*	-.41	.12	-.68*
4.no. identifications made				-	.56	.76*	.45	.73*	-.41	.12	-.68*
5. average confidence					-	.34	.52	.54	-.01	.48	-.32
6. estimated no.						-	.14	.84**	.11	.07	.02
7. conf. in 6.							-	.30	.45	.34	-.52
8. identify in court								-	.04	.17	.30
9. no. identifications made									-	.48	.68*
10. average confidence										-	-.09
11. hypnotic susceptibility											-

Note:-
*p<.05, **p<.01, ***p<.001.
Items 1 - 5 are before cognitive interview instructions. Items 6 - 10 are after cognitive interview instructions.

Table 18.3.1.11. Correlations (r) between performance variables for the verified group in the hypnosis condition.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. estimated no.	-	.19	.79*	.79**	.81*	.79*	.40	.50	.17	.48	.45	.46	.55	.79*	.56	.36	.28	.37	.57
2. conf. in 1.		-	.39	.11	.47	.31	-.20	.40	.32	.54	.26	.36	.17	.47	.27	-.07	.35	.32	-.35
3. identify in court			-	.79*	.90**	.69	.51	.44	-.12	.24	.71*	.78*	.08	.69	.16	-.08	.07	.21	.27
4.no. identifications made				-	.87**	.82**	.71*	.35	-.18	.50	.28	.61	.42	.81**	.38	.36	.02	.19	.47
5. average confidence					-	.88**	.39	.53	.16	.54	.80*	.77*	.39	.92**	.48	.12	.27	.47	.39
6. no. correct identifications (initial)						-	.18	.74*	.39	.74*	.53	.49	.58	.93***	.70*	.18	.53	.71*	.56
7. no. incorrect identifications (initial)							-	-.35	-.79	-.05	-.17	.45	.00	.25	-.22	.41	-.61	-.55	.12
8. no. correct identifications (total)								-	.71*	.58	.39	.20	.35	.61	.62	-.22	.75*	.88**	.50
9. accuracy rate (initial)									-	.49	.46	-.23	.44	.32	.67*	-.10	.83**	.87**	.24
10. estimated no.										-	.14	.14	.74*	.75*	.80*	.40	.65	.64	.17
11. conf. in 10.											-	.67	.14	.49	.31	-.21	.41	.51	.21
12. identify in court												-	.22	.62	.22	.15	.02	.14	.26
13. no. identifications made													-	.67*	.94***	.79*	.57	.55	.55
14. average confidence														-	.72*	.37	.40	.60	.52
15. no. correct identifications (initial)															-	.54	.78*	.80*	.57
16. no. incorrect identifications (initial)																-	.01	-.07	.34
17. no. correct identifications (total)																	-	.89**	.24
18. accuracy rate (initial)																		-	.48
19. hypnotic susceptibility																			-

Note:- *$p<.05$, **$p<.01$, ***$p<.001$. *Items 1 - 9 are before hypnosis. Items 10 -18 are after hypnosis.*

Table 18.3.1.12. Correlations (r) between performance variables for the unverified group in the hypnosis condition.

Measure	1	2	3	4	5	6	7	8	9	10	11
1. estimated no.	-	-.09	.73*	-.39	.42	.61	.14	.70	-.12	.45	.34
2. conf. in 1.		-	.26	-.36	.50	.56	.36	.30	-.20	.56	.17
3. identify in court			-	.02	.59	.89**	.25	.99***	-.17	.70	.22
4.no. identifications made				-	.06	-.19	.12	.03	.22	.22	-.08
5. average confidence					-	.57	.77**	.61	.53	.97***	-.03
6. estimated no.						-	.20	.89**	-.18	.72*	.51
7. conf. in 6.							-	.28	.19	.72*	-.14
8. identify in court								-	-.16	.73	.21
9. no. identifications made									-	.53	-.03
10. average confidence										-	.09
11. hypnotic susceptibility											-

Note:-
*$p<.05$, **$p<.01$, ***$p<.001$.
Items 1 - 5 are before hypnosis. Items 6 - 10 are after hypnosis.

Table 18.3.1.13. Correlations (r) between performance variables for the verified group in the control, cognitive interview and hypnosis conditions combined.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. estimated no.	-	.01	$.78^{***}$	$.63^{***}$	$.74^{***}$	$.67^{***}$	.35	$.54^{**}$	-.22	$.78^{***}$	.04	$.75^{***}$	$.50^{**}$	$.65^{***}$	$.48^{*}$	$.39^{*}$	$.43^{*}$	-.29	$.51^{*}$
2. conf. in 1.		-	.25	-.04	.12	.01	-.08	.04	.09	.04	.38	.13	-.18	.10	-.07	-.24	-.01	.15	-.29
3. identify in court			-	$.63^{***}$	$.80^{***}$	$.65^{***}$	.37	$.46^{*}$	-.13	$.61^{***}$	.26	$.82^{***}$	.25	$.62^{***}$	.27	.17	.23	-.10	.27
4.no. identifications made				-	$.73^{***}$	$.81^{***}$	$.81^{***}$	$.48^{*}$	$-.44^{*}$	$.45^{*}$	.08	.52	$.60^{**}$	$.59^{*}$	$.50^{**}$	$.53^{**}$	.31	-.34	$.40^{*}$
5. average confidence					-	$.80^{***}$	.39	$.57^{**}$	-.15	$.72^{***}$	.33	$.80^{***}$	$.53^{**}$	$.90^{***}$	$.54^{**}$	.38	$.44^{*}$	-.22	$.49^{*}$
6. no. correct identifications (initial)						-	.32	$.72^{***}$	.05	$.58^{**}$	.22	$.56^{**}$	$.56^{**}$	$.67^{***}$	$.71^{***}$	.27	$.54^{**}$	.03	$.49^{**}$
7. no. incorrect id (initial)							-	.06	$-.77^{***}$	.16	-.09	.28	$.41^{*}$	.30	.10	$.59^{**}$	-.04	$-.58^{**}$	.16
8. no. correct id (total)								-	-.06	$.54^{*}$	.17	.36	.38	$.59^{**}$	$.54^{**}$	.14	$.69^{***}$	-.12	$.45^{*}$
9. accuracy rate (initial)									-	-.10	.14	-.15	-.27	-.20	.10	$-.54^{**}$	-.06	$.87^{***}$	-.05
10. estimated no.										-	-.06	$.67^{***}$	$.56^{*}$	$.78^{***}$	$.54^{**}$	$.42^{*}$	$.53^{**}$	-.25	$.40^{*}$
11. conf. in 10.											-	.24	-.13	.27	.09	-.31	.20	.21	.04
12. identify in court												-	.37	$.69^{***}$	.39	.25	.36	-.15	$.43^{*}$
13. no. identifications made													-	$.67^{***}$	$.85^{***}$	$.87^{***}$	$.61^{***}$	$-.48^{*}$	$.62^{***}$
14. average confidence														-	$.65^{***}$	$.50^{**}$	$.56^{**}$	-.32	$.60^{**}$
15. no. correct identifications (initial)															-	$.48^{*}$	$.79^{***}$	-.04	$.63^{***}$
16. no. incorrect identifications (initial)																-	.28	$-.76^{***}$	$.44^{*}$
17. no. correct identifications (total)																	-	-.18	$.43^{*}$
18. accuracy rate (initial)																		-	-.20
19. hypnotic susceptibility																			-

Note:- $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$. *Items 1 - 9 are before interview instructions. Items 10 - 18 are after interview instructions.*

Table 18.3.1.14. Correlations (r) between performance variables for the unverified group for interview techniques combined.

Measure	1	2	3	4	5	6	7	8	9	10	11
1. estimated no.	-	-.09	.55**	.01	.44*	.61*	-.02	.53**	-.06	.41*	.20
2. conf. in 1.		-	.34	-.09	.24	.21	.29	.16	-.12	.27	.14
3. identify in court			-	.10	.42*	.53**	.23	.88***	-.09	.47*	.04
4.no. identifications made				-	.42*	.14	.29	.33	.51**	.29	-.16
5. average confidence					-	.35	.50**	.52**	.36	.75***	-.11
6. estimated no.						-	.08	.63***	-.12	.44	.25
7. conf. in 6.							-	.26	.04	.24	-.25
8. identify in court								-	.05	.48*	-.04
9. no. identifications made									-	.41*	-.32
10. average confidence										-	.05
11. hypnotic susceptibility											-

Note:-
$^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$.
Items 1 - 5 are before interview instructions. Items 6 - 10 are after interview instructions.

18.3.1.2 Actual performance of verified group

Further analysis, considering actual performance, was possible for the verified group. This was divided into three sub-sections 1) comparisons between subjects' estimates of their accuracy and their number of correct identifications, and 2), confidence-accuracy relationships and 3) effects of hypnotisability on 1 and 2.

Actual performance, was calculated with regard to number of correct identifications, number of incorrect identifications and accuracy rate. The means and SDs for the various analyses are displayed in Table 18.3.1.2.1.

Separate analysis was conducted on correct, incorrect and accuracy rate for initial identifications and total identifications. Hypnotic susceptibility was considered with Pearson's correlations which are reported later. It was not included as a factor in these ANOVAs because the limited number of subjects in each interview condition reduced cell size to such an extent as to make analysis impracticable.

A 3 X 2 ANOVA with repeated measures on the second factor (control/cognitive interview/hypnosis X before/after) on correct initial identifications showed no significant main effect of interview condition, $F(2,23)=1.18$ $p>.32$. However, there was a significant increase in the number of correct identifications that were made after interview instructions $F(1,23)=6.16$, $p<.02$. The interaction between interview condition and before/after interview instructions was not significant, $F(2,23)=2.34$ $p>.12$.

A 3 X 2 ANOVA with repeated measures on the second factor (control/cognitive interview/hypnosis X before/after) on initial incorrect identifications showed no significant main effect of interview condition, $F(2,23)=2.27$ $p>.13$. However, there was a significant increase in the number of incorrect identifications after interview instructions $F(1,23)=5.66$, $p<.03$. The interaction between interview condition and before/after interview instructions was not significant, $F(2,23)=2.96$ $p>.07$.

A 3 X 2 ANOVA with repeated measures on the second factor (control/cognitive interview/hypnosis X before/after) on initial accuracy rate showed no significant main effect of interview condition, $F(2,23)=0.72$ $p>.50$ or of before/after interview instructions $F(1,23)=2.40$, $p>.13$. The interaction between interview condition and before/after interview instructions was not significant, $F(2,23)=2.96$, $p>.07$.

When total identifications were considered a 3 X 2 ANOVA with repeated measures on the second factor (control/cognitive interview/hypnosis X before/after) no significant main effect was found of interview condition on the overall number of correct identifications, $F(2,22)=0.25$ $p>.78$, or before/after interview instructions $F(1,22)=0.00$, $p>.97$. No significant interaction was found. Because both the total number of incorrect identifications and the total accuracy rate were related to the total number of correct identifications by the nature of the forced-recall task further analyses of total incorrect and accuracy rate were unnecessary.

Table 18.3.1.2.1.1 Control, cognitive interview and hypnosis performance before and after procedure considering actual performance.

measure	control N=9		cognitive interview N=8		hypnosis N=9		combined N=26	
	before	after	before	after	before	after	before	after
initial identifications								
no. of correct identifications	12.33 (2.34)	14.05 (2.55)	11.00 (4.90)	14.05 (2.55)	9.33 (4.36)	12.11 (5.04)	12.08 (4.02)	12.36 (4.19)
no. of incorrect identifications	8.67 (4.27)	7.09 (4.06)	4.00 (2.00)	6.62 (4.03)	4.67 (3.32)	8.22 (2.86)	5.84 (3.86)	7.64 (3.60)
accuracy rate	.61 (.13)	.66 (.16)	.73 (.12)	.64 (.09)	.68 (.16)	.64 (.09)	.67 (.13)	.63 (.12)
total identifications								
no. of correct identifications	13.44 (2.01)	14.11 (2.32)	14.63 (2.56)	13.75 (2,25)	13.25 (3.28)	13.50 (2.51)	13.76 (2.60)	13.88 (2.27)
no. of incorrect identifications	11.56 (2.01)	10.89 (2.32)	10.38 (2.56)	10.25 (2.25)	11.75 (3.28)	11.50 (2.51)	11.24 (2.60)	11.12 (2.27)
accuracy rate	.54 (.08)	.56 (.09)	.59 (.10)	.55 (.09)	.53 (.13)	.54 (.10)	.55 (.10)	.56 (.09)
confidence in correct identifications	2.72 (0.81)	2.83 (0.85)	2.85 (0.79)	3.02 (0.76)	2.87 (0.87)	3.48 (1.01)	2.81 (0.79)	3.12 (0.87)
confidence in incorrect identifications	1.81 (0.56)	1.96 (0.73)	1.55 (0.57)	1.79 (0.78)	1.69 (0.71)	2.37 (0.85)	1.69 (0.60)	2.04 (0.77)
confidence in correct/incorrect identifications (combined means)	2.27 (0.62)	2.40 (0.72)	2.20 (0.63)	2.40 (0.71)	2.28 (0.75)	2.93 (0.87)	2.25 (0.64)	2.58 (0.76)

Note: Standard deviations in brackets.

Subjects' estimates of how many accurate identifications they could make both before and after interview instructions and the actual number of correct identifications that subjects made were compared with a 3 X 2 X 2 ANOVA with repeated measures on the second and third factors (control/cognitive interview/hypnosis X subjects' estimates/no. of correct identifications X before/after). No significant effect was found of interview condition $F(2,22)=0.06$ $p>.94$. There was no significant before/after effect, $F(1,22)=0.05$ $p>.82$. Subjects' estimates of how many accurate identifications they could make was not significantly different from the number of accurate identifications that they did make, $F(1,22)=0.04$ $p>.84$. A significant interaction was found between subjects' estimates of how many accurate identifications they could make/number correct identifications that they made and before/after interview instructions, $F(2,22)=5.95$ $p<.03$. Follow-up *F*-tests did not show any differences between any of these cells. No significant interactions were found between interview condition and subjects' estimates of how many accurate identifications they could make/number correct identifications that they made, interview condition and before/after interview instructions, and between interview condition and subjects' estimates of how many accurate identifications they could make/number correct identifications that they made and before/after interview instructions, $F(2,22)=2.09$ $p>.15$, $F(2,22)=2.96$ $p>.07$, and $F(2,22)=0.41$ $p>.67$, respectively.

The number of identifications that subjects stated they would identify in court both before and after interview instructions and the actual number of correct identifications that subjects made were compared with a 3 X 2 X 2 ANOVA with repeated measures on the second and third factors (control/cognitive interview/hypnosis X no. identify in court/no. of correct identifications X before/after). No significant effect was found of interview condition $F(2,22)=0.64$ $p>.54$. There was no significant before/after interview instructions effect on the number of identifications that subjects stated that they could identify in court and the number of correct identifications that they made $F(1,22)=3.18$ $p>.08$. The number of identifications that subjects stated that they would make in court was significantly less than the number of correct identifications that they did make, $F(1,22)=23.99$ $p>.0001$. No significant interactions were found between interview condition and before/after interview instructions, interview condition and the number of identifications that subjects would identify in court/number correct identifications that they made, the number of identifications that subjects would identify in court/number correct identifications that they made and before/after interview instructions, and between interview condition and subjects' estimates of how many accurate identifications they could make/number correct identifications that they made and before/after interview instructions $F(2,22)=1.27$ $p>.30$, $F(2,22)=0.24$ $p>.79$, $F(2,22)=0.09$ $p>.77$ and $F(2,22)=1.70$ $p>.21$, respectively.

A 3 X 2 X 2 ANOVA with repeated measures on the second and third factors (control/cognitive interview/hypnosis X subjects' estimates of the number of photographs they could accurately identify/the number of identifications subjects stated that they would identify in court X before/after interview instructions) indicated that there was no significant effect of interview condition, $F(2,22)=0.01$ $p>.99$. Subjects again made significantly greater estimates of the number of photographs that they could accurately identify than the number that they said they could identify in court, $F(2,22)=40.00$

$p < .0001$. There was no significant before/after effect $F(2,22)=0.02$ $p > .89$. No significant interactions were found between, interview instructions and subjects' estimates of the number of photographs they could accurately identify/the number of identifications subjects stated that they would identify in court, interview condition and before/after interview instructions, before/after interview instructions and subjects' estimates of the number of photographs they could accurately identify/the number of identifications subjects stated that they would identify in court and between interview condition and before/after interview instructions and subjects' estimates of the number of photographs they could accurately identify/the number of identifications subjects stated that they would identify in court $F(2,22)=1.14$ $p > .34$, $F(2,22)=2.34$ $p > .12$, $F(2,22)=3.16$ $p > .09$, $F(2,22)=0.56$ $p > .58$.

Pearson's correlations were conducted between a variety of subjects' performance measures and their actual accuracy in terms of correct identifications, incorrect identifications and accuracy rate for each interview condition (all are shown in the previous tables). Because these correlations were broadly similar for each interview condition and so many correlations were calculated, only the combined correlations of the control, cognitive interview and hypnosis conditions results are described here. Generally, both before and after interview instructions the following performance measures were significantly correlated with one-another; subjects's estimates of their performance (how many photographs they could accurately identify; how many photographs they could identify in court) the number of identifications that subjects made, subjects' average confidence in identifications, the number of initial correct identifications and the number of total correct identifications.

Accuracy rate was significantly negatively correlated with the number of incorrect identifications before and after interview instructions. Of note was that the number of incorrect identifications was not significantly correlated with the number of correct identifications before interview instructions although it was after interview instructions.

When hypnotic susceptibility was considered, for all three conditions combined, a significant correlation was found between hypnotisability and the number of correct initial and total identifications before interview instructions. After interview instructions hypnotic susceptibility was significantly correlated with the number of correct initial and total identifications before interview instructions and also the number of incorrect identifications. Interestingly, if these correlations are considered for each condition separately, if anything, they were stronger in the cognitive interview and control conditions than the hypnosis condition.

Within-subjects C-A correlations were calculated for each subject in each interview condition before and after interview instructions. These correlations were converted to z-scores, averaged and tested for significance against zero, results are shown in Table 18.3.1.2.2. These correlations were significant for each interview condition before and after interview instructions.

A 3 X 2 ANOVA with repeated-measures on the second factor (control/cognitive interview/hypnosis X before/after interview instructions) indicated no significant main effect of interview condition, $F(2,22)=2.82$ $p>.08$ or of before/after interview instructions, $F(1,22)=0.28$ $p>.60$, on these z scores. No interaction was found, $F(2,22)=0.21$ $p>.81$.

Table 18.3.1.2.2. Within subjects C-A correlations for the face recognition task.

	interview condition			
	control	cognitive interview	hypnosis	combined
before	.47*	.47*	.48*	.47*
	(.16)	(.16)	(.17)	(.15)
	N=9	N=8	N=8	N=25
after	.43*	.47*	.48*	.46*
	(.21)	(.12)	(.23)	(.18)

Note. Standard deviations are in brackets. * indicates *z*-score significantly different from zero $p<.05$, ** indicates $p<.01$, *** indicates $p<.001$.

Each subjects' average accuracy (i.e. mean number of correct identifications) was correlated with his/her average confidence rating in those identifications. The results are shown in Table 18.3.1.2.3. Correlations for the control, cognitive interview and hypnosis conditions combined were significant before and after interview instructions. Also, correlations for the cognitive interview group before interview instructions and the control group after instructions were significant.

Table 18.3.1.2.3. Subjects' average confidence correlated with subjects' average accuracy for identifications with respect to interview conditions.

	interview condition			
	control	cognitive interview	hypnosis	combined
before interview instructions	.37	.84**	.53	.57**
after interview instructions	.88**	.56	.40	.57**

Note. *indicates $p < .05$, **indicates $p < .01$, *** indicates $p < .001$.

A between subjects C-A correlation was calculated for each identification that subjects made (i.e. C-A correlations were calculated across subjects for the first identification that subjects made, a C-A correlation was calculated across subjects for the second identification that subjects made, etc. and these correlations were averaged). None was significant. These are displayed in Table 18.3.1.2.4.

Table 18.3.1.2.4. Between subjects C-A correlations for the face recognition task.

	interview condition			
	control	cognitive interview	hypnosis	combined
before	.22 N=25	.26 N=25	.23 N=23	.27 N=25
after	.27 N=25	.24 N=25	.25 N=24	.25 N=25

Note. Standard deviations are in brackets. * indicates *z*-score significantly different from zero $p<.05$, ** indicates $p<.01$, *** indicates $p<.001$.

In addition, the average accuracy and average confidence score for each identification number was correlated (i.e. the average confidence expressed in the first identification and the average accuracy of the first identification, the average confidence expressed in the second identification and the average accuracy of the second identification etc. were correlated). These correlations were significant for each interview condition before and after interview instructions and are displayed in Table 18.3.1.2.4.

Table 18.3.1.2.4. Average confidence expressed in each identification number correlated with their average accuracy for that number identification with respect to interview condition for identifications.

	interview condition			
	control	cognitive interview	hypnosis	combined
before interview instructions	.78***	.75***	.84***	.90***
after interview instructions	.66***	.76***	.71***	.91***

Note. *indicates $p < .05$, **indicates $p < .01$, *** indicates $p < .001$.

Confidence ratings in correct and incorrect identifications are displayed in Table 18.3.1.2.1. A 3 X 2 X 2 ANOVA with repeated measures on the second and third factors (control/cognitive interview/hypnosis X correct/incorrect identifications X before/after interview instructions) was conducted on these confidence scores. No significant main effect was found for interview condition $F(2,22)=0.45$ $p > .64$. However, a significant main effect was found with regards to confidence in correct/incorrect identifications; greater confidence was expressed in accurate identifications $F(1,22)=91.86$ $p > .0001$. A significant main effect of before/after interview instructions was found $F(1,22)=31.72$ $p < .0001$ indicating that confidence in both correct and incorrect identifications increased after interview instructions. There was a significant interaction between interview condition and before/after interview instructions $F(2,22)=7.70$ $p<.0029$. Neither the interaction between interview condition and correct/incorrect identifications or between interview condition and correct/incorrect identifications and before/after interview instructions were significant, $F(2,22)=0.92$ $p<.41$ and $F(2,22)=0.01$ $p>.98$ respectively.

To investigate the significant interaction further, follow-up F tests were conducted. No differences were found between interview conditions before and after interview instructions. However, both the hypnosis group and the cognitive interview group increased their combined confidence in correct/incorrect identifications after interview instructions.

Further investigation of this increase was possible by creating a new variable by subtracting confidence expressed in correct and incorrect identifications before interview instructions from confidence expressed in correct and incorrect identifications after interview instructions. A one-way ANOVA (control/cognitive interview/hypnosis) was conducted on this variable that showed a significant effect of interview condition, $F(2,22)=7.70$ $p<.0030$. Follow-up F tests indicated that the hypnosis group expressed significantly greater confidence than the control and cognitive interview group. The control group and cognitive interview group did not significantly differ from one another.

When 'absolutely certain' identifications were considered independently a Chi-square analysis showed no significant differences for interview condition for the number of correct or incorrect absolutely certain responses. Of the identifications that subjects, all conditions combined, rated as being absolutely certain that they were correct 84.98% were correct.

Table 18.3.1.2.5. Breakdown of 'Absolutely certain' responses with respect to interview condition and accuracy.

	control	cognitive interview	hypnosis	combined
before interview instructions				
correct	26	39	41	106
incorrect	8	4	6	18
after interview instructions				
correct	33	39	54	126
incorrect	4	6	13	23
overall				
correct	59	78	95	232
incorrect	12	10	19	41

Pearson's product moment correlations were calculated with hypnotic susceptibility and the following variables: confidence in correct answers, confidence in incorrect answers and within-subjects C-A correlations, number of identifications made and average confidence expressed in identifications both before and after interview instructions. The results are displayed in Table 20.

Hypnotisability was significantly correlated with the confidence expressed in correct identifications both before and after interview instructions when interview conditions were combined. Together with the previously reported findings, this indicates that subjects that reported greater hypnotisability were not only more likely to make more accurate identifications, but also to express greater confidence in accurate identifications. Higher hypnotisability was also significantly correlated with higher confidence in correct identifications in the control group before/after interview instructions and the cognitive interview group with higher numbers of identifications after interview instructions. However, it should be noted that some non-significant correlations were quite high, for example between hypnotisability and the number of correct identifications in the cognitive interview group (r=.69); however, small N's prevented many of these correlations reaching significance.

Table 18.3.1.2.3.1 Correlations of hypnotisability with identification performance for control, cognitive interview and hypnosis conditions before and after interview procedure.

measure	control		cognitive interview		hypnosis		combined	
	before	after	before	after	before	after	before	after
confidence in correct identification	.73*	.69*	.33	.50	.38	.67	.46*	.65***
confidence in incorrect identification	.45	.45	.40	.14	.24	.22	.34	.33
within-subjects C-A relationships	.16	.04	-.10	.36	.17	.56	.10	.34

Note: Standard deviations in brackets.

18.3.2. Questionnaire recall task

The recall questionnaire was analyzed in a similar manner to that of Experiment 7 and the face recognition questionnaire. Hypnotic susceptibility was originally included in the ANOVAs outlined below. However, it was subsequently dropped because no significant effects were found. Instead the influence of hypnotic susceptibility was investigated with Pearson's correlations.

A 3 X 3 mixed ANOVA with repeated measures on the second factor (control/cognitive interview/hypnosis X easy/medium/hard questions) was conducted for correct answers. The means and standard deviations are displayed in Table 18.3.2.1. The main effect for Interview condition was not significant, $F(2,47)=0.53$, $p>.59$; however, a significant main effect was found for question difficulty $F(2,48)=748.92$, $p<.0001$. Follow-up F tests ($p<.05$) confirmed that these differences were in the appropriate direction; easy questions were significantly more likely to be answered correctly than medium questions, which in turn were significantly more likely to be answered correctly than hard questions.

The interaction between interview condition and question difficulty was also significant, $F(4,47)=2.47$, $p<.05$. Follow-up F tests ($p<.05$) indicated that there were no significant differences between conditions when easy, medium or hard questions were considered independently. So to explore this interaction further three 3 X 2 ANOVAs were conducted, for interview condition by easy and medium questions, interview condition by medium and hard questions, and interview condition by easy and hard questions. The only significant interaction was for easy and medium question X interview condition,

$F(2,48) = 3.69$, $p < .032$. Further F tests showed each interview condition had significantly more correct answers for easy questions than medium questions. Therefore, a one-way ANOVA (by interview conditions) was conducted on the number of easy questions that were answered correctly subtracted from the number of medium questions answered correctly. The main effect for conditions was significant, $F(2,48) = 3.69$ $p < .04$, and follow-up F tests indicated that the was the difference between the number of easy and medium questions answered correctly was greater in the control group than in the hypnosis group ($p < .05$).

Table 18.3.2.1. Number of correct answers with respect to interview condition and question difficulty.

question difficulty	interview condition			
	control	cognitive interview	hypnosis	combined
easy	10.19 (0.75)	9.94 (1.09)	9.89 (0.83)	10.00 (0.89)
medium	4.25 (1.69)	5.29 (1.93)	5.44 (1.98)	5.02 (1.91)
hard	1.00 (0.97)	0.65 (0.61)	1.06 (1.16)	0.90 (0.94)
overall	15.43 (2.55)	15.88 (2.78)	16.38 (2.70)	15.92 (2.66)

Note. Standard deviations are in brackets.

A C-A correlation was calculated for each subject across the 33 questions, producing 51 correlations. As outlined in previous experiments a *z* score was calculated for each subject's correlation, and the average of these *z* scores was tested for significance against zero. Further within-subjects correlations were calculated for the different permutations of interview condition and question difficulty. These C-A correlations were also transformed into *z*-scores and the average *z* score tested against zero. The average of these various correlations and their levels of significance (based on the converted *z* scores) are shown in Table 18.3.2.2. As can be seen in Table 18.3.2.2, all the within-subjects C-A correlations are significant with the exception of those for the hard questions alone, and for the medium questions in the control group. The correlation for medium questions for the control group (r=.69) was not significant while the cognitive interview correlation was actually less (r=.68) but was significant. This due to the smaller number of subjects in the control condition compared to the cognitive interview group.

To assess the effects of interview condition and question difficulty on within-subjects C-A relations, ideally a 3 X 2 ANOVA with repeated measures as the second factor (control/cognitive interview/hypnosis X easy/medium/hard questions) would have been conducted on the *z* scores. However, this was not possible for this experiment as a large number of subjects either answered all the easy questions correctly or answered all the difficult questions incorrectly so correlations could not be calculated for them and therefore they would have to be dropped from the analysis leaving cell sizes too small to conduct an analysis.

Instead, a series of one-way ANOVAs were conducted. First, overall *z* scores for the C-A relationship for each subject were compared across interview condition (control/cognitive interview/hypnosis). The main effect for conditions was not significant, $F(2,48)=2.56$, $p>.09$. The main effect for question difficulty (conditions combined) was not significant, $F(2,13)=2.36$, $p>.13$, although it should be noted that the number of subjects that could be used in this analysis was only eight. Further analyses were then conducted on easy, medium and hard questions separately. There was a significant effect of condition on easy questions, $F(2,11)=4.18$ $p<.05$. Follow up F-tests indicated that the mean C-A correlation for the cognitive interview was significantly greater than that for the control group ($p<.05$). No significant difference was found between the control group and the cognitive interview group or the cognitive interview group and the hypnosis group. There were no effects of interview condition on medium or hard questions, $F(2,48)=0.34$ $p>.71$ and $F(2,28)=1.00$ $p>.38$ respectively.

Table 18.3.2.2. Within-subjects C-A correlations with respect to interview condition and question difficulty.

question difficulty	interview condition			
	control	cognitive interview	hypnosis	combined
easy	.86** (.14) N=7	.96* (.07) N=7	.86* (.10) N=7	.89** (.11) N=21
medium	.69 (.20) N=16	.68** (.22) N=18	.68* (.14) N=17	.69* (.19) N=51
hard	.62 (.38) N=11	.47 (.51) N=10	.45 (.36) N=10	.52 (.41) N=31
overall	.84** (.08) N=16	.83** (.06) N=17	.79* (0.08) N=18	.82** (.08) N=51

Note. Standard deviations are in brackets. * indicates *z*-score significantly different from zero $p<.05$, ** indicates $p<.01$, *** indicates $p<.001$.

Each subjects' average accuracy was correlated with his/her average confidence rating. These correlations were further broken down into question difficulty and interview condition. The results are shown in Table 18.3.2.3. All the correlations are significant except for those in the hard category.

Table 18.3.2.3. Subjects' average confidence scores correlated with their average accuracy with respect to interview condition and question difficulty.

question difficulty	interview condition			
	control	cognitive interview	hypnosis	combined
easy	.78*** N=16	.88*** N=17	.73*** N=18	.75*** N=51
medium	.67** N=16	.60* N=17	.62** N=18	.62* N=51
hard	.39 N=16	.29 N=17	.04 N=18	.25 N=51
overall	.69** N=16	.78** N=17	.63** N=18	.68*** N=51

Note. Standard deviations are in brackets. *indicates $p<.05$, **indicates $p<.01$, *** indicates $p<.001$.

The C-A relationship was then calculated across-subjects for each of the 33 questions. These correlations were averaged, converted to z-scores and tested for significance against zero. Further correlations and z scores were calculated for the various different interview conditions and question difficulties. These correlations and significance levels for the z scores are shown in Table 18.3.2.4. Only one was significant; that for easy questions in the control group; however, it can be noted that a number of the N's in this particular analysis were extremely small. Given these findings no further analyses were conducted on these correlations.

Table 18.3.2.4. Between-subjects C-A correlations for interview condition and question difficulty.

question difficulty	interview condition			
	control	cognitive interview	hypnosis	combined
easy	.93*	.77	.58	.73
	(.05)	(.36)	(.42)	(.35)
	N=4	N=5	N=6	N=15
medium	.45	.50	.52	.49
	(.42)	(.45)	(.39)	(.41)
	N=11	N=10	N=11	N=32
hard	.63	.30	.51	.51
	(.33)	(.51)	(.32)	(.35)
	N=5	N=3	N=7	N=15
overall	.59	.54	.53	.54
	(.39)	(.44)	(.36)	(.39)
	N=20	N=18	N=24	61

Note. Standard deviations are in brackets. * indicates *z*-score significantly different from zero $p<.05$, ** indicates $p<.01$, *** indicates $p<.001$.

The average number of correct answers and average confidence score for each question was correlated. This procedure was again also conducted for the various permutations of interview condition and item difficulty; the results are displayed in Table 18.3.2.5. All these correlations were significant.

Table 18.3.2.5. Questions' average confidence correlated with their average accuracy with respect to interview condition and question difficulty.

question difficulty	interview condition			
	control	cognitive interview	hypnosis	combined
easy	.95*** N=11	.96*** N=11	.88*** N=11	.97*** N=11
medium	.79** N=11	.75** N=11	.84** N=11	.81** N=11
hard	.88*** N=11	.82** N=11	.55 N=11	.73* N=11
overall	.97*** N=33	.96*** N=33	.96*** N=33	.97*** N=33

Note. *indicates $p<.05$, **indicates $p<.01$, *** indicates $p<.001$.

Subjects' confidence in correct answers was calculated for easy medium and difficult questions (see Table 18.3.2.6). A 3 X 3 mixed ANOVA with repeated measures on the second factor (control/cognitive interview/hypnosis X easy/medium/hard questions) was conducted on subjects' confidence in correct answers. A significant effect was found of question difficulty on confidence in correct answers $F(2,27)=41.10$, $p<.0001$. Follow-up F tests indicated that subjects expressed significantly higher confidence in easy questions that were answered correctly than medium questions that were answered correctly which in turn had higher confidence ratings than hard questions. No significant main effect was found of interview condition $F(2,27)=1.13$, $p>.33$ and no significant interaction was found between condition and question difficulty $F(2,27)=0.28$, $p>.76$. As some subjects answered all the hard questions incorrectly they could not be used in the analysis, means displayed in bold in Table 18.3.2.6 are those used for the above ANOVA.

Table 18.3.2.6. Average confidence expressed in correct answers with respect to interview condition and question difficulty.

question difficulty	interview condition			
	control	cognitive interview	hypnosis	combined
easy	9.40	9.52	8.97	9.29
	(0.73)	(0.45)	(1.23)	(0.89)
	N=16	N=17	N=18	N=51
	9.77	**9.69**	**9.08**	**9.54**
	(0.30)	**(0.20)**	**(0.75)**	**(0.54)**
	N=11	**N=10**	**N=9**	**N=30**
medium	6.92	6.80	6.77	6.83
	(1.54)	(1.80)	(1.84)	(1.70)
	N=16	N=17	N=18	N=51
	6.74	**7.17**	**6.95**	**6.95**
	(1.57)	**(1.71)**	**(2.20)**	**(1.77)**
	N=11	**N=10**	**N=9**	**N=30**
hard	6.21	4.85	3.63	4.98
	(3.63)	(3.59)	(2.61)	(3.40)
	N=11	N=10	N=9	N=30
	6.21	**4.85**	**3.63**	**4.98**
	(3.63)	**(3.59)**	**(2.61)**	**(3.40)**
	N=11	**N=10**	**N=9**	**N=30**
overall	7.92	7.78	7.35	7.68
	(0.60)	(0.88)	(1.27)	(0.97)
	N=16	N=17	N=18	N=51
	8.66	**8.58**	**7.86**	**8.39**
	(0.51)	**(0.70)**	**(1.22)**	**(0.88)**
	N=11	**N=10**	**N=18**	**N=30**

Note. Standard deviations are in brackets, means in bold were used for the analysis of variance.

Subjects' average confidence in incorrect answers was calculated (see Table 18.3.2.7). A 3 X 3 mixed ANOVA with repeated measures on the second factor (control/cognitive interview/hypnosis X easy/medium/hard questions) was conducted on subjects' confidence in incorrect answers. A significant effect for question difficulty on confidence in incorrect answers was found $F(2,32)=7.33$, $p<.002$. Follow-up F tests indicated that subjects gave significantly higher confidence scores for medium questions that were answered incorrectly than for hard questions that were answered incorrectly ($p<.05$); the other comparisons were not significant. No main effect was found for interview condition, $F(2,32)=1.28$, $p>.29$ or interaction between question difficulty and interview condition, $F(4,32)=0.42$, $p>.79$. As some subjects answered all the easy questions incorrectly they could not be used in the analysis, means displayed in bold in Table 18.3.2.7 are those used for the above ANOVA.

Table 18.3.2.7. Average confidence expressed in incorrect answers with respect to interview condition and question difficulty.

question difficulty	interview condition			
	control	cognitive interview	hypnosis	combined
easy	2.25	1.75	2.80	2.34
	(1.36)	(1.56)	(2.98)	(1.22)
	N=10	N=10	N=15	N=35
	2.25	**1.75**	**2.80**	**2.34**
	(1.37)	**(1.55)**	**(2.98)**	**(2.23)**
	N=10	**N=10**	**N=15**	**N=30**
medium	3.19	2.67	3.01	2.95
	(1.35)	(1.23)	(1.10)	(1.22)
	N=16	N=17	N=18	N=51
	3.04	**2.40**	**2.90**	**2.80**
	(1.28)	**(1.14)**	**(1.13)**	**(1.17)**
	N=10	**N=10**	**N=10**	**N=30**
hard	1.71	1.39	1.57	1.55
	(0.67)	(0.43)	(0.38)	(0.51)
	N=16	N=17	N=18	N=51
	1.82	**1.22**	**1.53**	**1.53**
	(0.81)	**(0.18)**	**(0.38)**	**(0.54)**
	N=10	**N=10**	**N=15**	**N=30**
overall	2.25	1.85	2.13	2.07
	(0.83)	(0.60)	(0.55)	(0.67)
	N=16	N=17	N=18	N=51
	2.31	**1.65**	**2.06**	**2.02**
	(0.98)	**(0.42)**	**(0.52)**	**(0.69)**
	N=10	**N=10**	**N=15**	**N=35**

Note. Standard deviations are in brackets. Means used for ANOVAs are in bold.

Confidence in correct answers was compared with confidence in incorrect answers with a 3 X 2 ANOVA (control/cognitive interview/hypnosis X confidence in correct/incorrect answers). This indicated that subjects showed higher confidence in correct answers than incorrect answers, $F(1,47)=1741.61$ $p<.0001$. There was no effect of interview condition, $F(2,47)=1.17$ $p>.32$ although there was a trend towards an interaction between confidence in correct/incorrect answers and interview condition, $F(2,47)=2.50$ $p<.10$.

Independent 3 X 2 ANOVAs (control/cognitive interview/hypnosis X confidence in correct/incorrect answers) were conducted easy, medium and difficult answers independently (to preserve cell size). Confidence in correct answers was significantly greater than confidence in incorrect answers for easy questions, $F(1,32)=332.18$ $p<.0001$. There was no effect of interview condition, $F(2,32)=0.05$ $p>.95$, and no significant interaction, $F(2,32)=1.96$ $p>.16$. Similar results were found for medium and hard questions. For medium questions confidence in correct answers was significantly higher than confidence in incorrect answers, $F(1,48)=279.90$ $p<.0001$. There was no effect of interview condition, $F(2,48)=0.26$ $p>.77$, or interaction, $F(2,48)=0.30$ $p>.74$. For hard questions confidence in correct answers was significantly higher than confidence in incorrect answers, $F(1,27)=32.68$ $p<.0001$. There was no effect of interview condition, $F(2,27)=1.39$ $p>.27$, or interaction, $F(2,27)=1.59$ $p>.22$.

Absolutely sure responses were considered with a Chi-square analysis. There was no significant effect of interview condition and correct/incorrect responses. The overall accuracy rate for absolutely sure responses was When accuracy rate was considered it was 98.54% overall. The breakdown of absolutely sure responses is displayed in Table 18.3.2.8.

Table 18.3.2.8. Breakdown of 'Absolutely certain' responses with respect to interview condition and correct responses.

	control	cognitive interview	hypnosis	combined
correct	138	174	159	472
incorrect	3	1	3	7
overall	141	175	162	479

No significant correlations were found between hypnotic susceptibility and the following variables: number of correct answers, number of incorrect answers, confidence in correct answers, confidence in incorrect answers and within-subjects C-A correlations. All these analyses were further broken down with respect to question difficulty, but again no significant correlations were found ($p > .10$ in all cases).

18.3.3. Comparisons between recall questionnaire and identification questionnaire

Correlational analysis was conducted between the questionnaire recall task and the face recognition task to determine if subjects' performance was related on the two tasks.

To allow comparability, as the recall task was undertaken after interview instructions it was compared with face-recognition performance also after interview instructions. Thus, for example, confidence in correct responses (answers) on the recall task were correlated with correct responses on the face-recognition task (identifications). The variables that were used for this analysis were the number or correct responses, confidence in correct responses, confidence in incorrect responses, average confidence in responses and within subjects C-A correlations. The results are displayed in Table 18.3.3.1.

Two correlations were significant, one for the control condition and one for the hypnosis condition, these were both between confidence in correct responses on the recall task and the face-recognition task. This indicates that subjects in the control condition and hypnosis condition who expressed high confidence in correct answers on the recall task also expressed high confidence in correct identifications on the face recognition task.

Table 18.3.3.1. Correlations between recognition and recall questionnaires.

measure	interview condition			
	control N=9	cognitive interview N=8	hypnosis N=9	combined N=26
no. correct	.56	.28	.56	.37
within C-A correlation	-.08	-.41	-.08	-.16
confidence in correct responses	.78*	.05	.78*	.28
confidence in incorrect responses	.05	.19	.05	.23

Note. Standard deviations are in brackets.

18.4 Discussion

18.4.1 Face recognition task

18.4.1.1 Verified/unverified group comparisons.

The initial hypotheses regarding verification were: 1) subjects' whose answers could not be verified would estimate that they could accurately identify significantly more photographs than subjects whose answers could be verified; 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimates of how many photographs that they could accurately identify, and the accuracy of their actual identifications, than subjects whose answers could be verified; and 3) subjects' whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court than subjects whose answers could be verified.

None of these hypotheses was supported. This would seem to suggest that verifiability does not, using this paradigm, affect eyewitness performance on these variables, even if procedures which may place considerable demands on the eyewitness, such as hypnosis or a cognitive interview are used.

However, some significant effects of interview instructions were found. The number of identifications that subjects made increased after interview instructions were given for all three interview conditions. This may have been due to the nature of the experimental conditions rather than the interview instructions themselves. Subjects initially filled in the answer sheet and if they had not made 25 identifications were instructed to try again. Thus, when they filled in the answer sheet the second time, after interview instructions, they may have made more identifications because they knew that they would be required to make 25. However, as the effect also occurred in some of the previous studies in which forced-choice testing was not used, it is likely that this effect was at least in part due to repeated testing (see Nogrady, McConkey and Perry, 1985).

Average confidence in identifications was also increased after interview instructions for all three interview conditions. However, an interaction was also found, subjects in the hypnosis condition increased their confidence more than the other conditions. This fits in well with what might have been expected on the basis of the literature (see, for example, Wagstaff, 1989, 1993, 1995). Hypnosis, may possess characteristics that set it apart from other interview procedures. For instance, there is an extra investment in length of time spent 'hypnotising' the witness, with the implicit expectation their performance will be enhanced; this might place additional pressure on the eyewitness to perform like a good subject and therefore to be confident. Also, given the public belief that hypnosis improves accurate testimony, witnesses who believe they are or have been 'hypnotised' may place greater credibility on their memories. Correlations between performance measures for the verified group and the unverified group were once again broadly similar. Generally, subjects who made high estimates of the number of photographs that they could identify appeared to do so consistently, whether the measure was their estimates of how many photographs they could accurately identify, the number they would identify in court and regardless of whether this was before/after interview instructions. These subjects also tended to make more identifications and express greater confidence in individual identifications. As such, these findings were similar to those of previous experiments.

However, some effects were only significant for the verified group. In particular, hypnotisability for the verified group was significantly correlated with subjects' estimates of how many accurate identifications the could make before and after interview instructions, the number of identifications that they would make in court after interview instructions, the number of identifications made before and after interview instructions and the average confidence in identifications before and after interview instructions. Broadly speaking, these effects appeared to occur across interview conditions not simply in the hypnosis condition as one might expect. None of these effects was significant for the unverified group.

It is not immediately obvious why this discrepancy should occur as there was no evidence, for example, that the unverified group inflated their estimates more than the verified group, and an ANOVA verified/unverified X high/low susceptibility, showed no significant interaction. One is, perhaps, left wondering whether these differences are reliable, given the large number of correlations.

18.4.1.2 Actual performance of verified group

It was hypothesised that within the verified group there would be a positive relationship between subjects' estimates of how many photographs they thought they could accurately identify and the number of correct identifications they made. This was supported both by comparisons between subjects' estimates of how many accurate identifications they could make before and after interview instructions and the actual number of correct identifications that subjects made after interview instructions and correlations between the two measures. Thus, subjects' estimates of how many photographs they could identify appears to be related to the number of correct identifications that they made. This is broadly consistent with previous experiments.

When the number of identifications that subjects stated they would identify in court before and after interview instructions were compared with the number of correct identifications that they made, significant differences were found. Although subjects increased the number of photographs that they stated that they would identify in court after interview instructions, once again this estimate was still significantly less than the number of photographs that they could identify. In addition, comparisons between subjects' estimates of the number of photographs they could accurately identify, and the number of identifications subjects stated that they would identify in court, before and after interview instructions, indicated that subjects made significantly higher estimates of the number of photographs that they could accurately identify than the number that they said they could identify in court. Thus, although the number of photographs that subjects stated that they would identify in court was increased by interview instructions, it was still a conservative estimate, compared to the actual number of correct identifications that subjects made. This finding was consistent with Experiments 2, 4 and 8.

Because before-after identifications and a forced-choice measures were taken in this experiment it was possible to determine whether these effects were due to real enhancements in recall or simply reflected a lowered criterion for report. Of particular interest, therefore, in this experiment, was the finding that both the number of (initial) correct and (initial) incorrect identifications were increased by interview instructions. Nevertheless, the ANOVAs for the forced-choice (total) correct identifications revealed no before/after effects whatsoever, regardless of interview instructions. Taken as a whole, this would suggest that if subjects are asked to repeat a memory task, they will lower their criterion for report. In real-life situations, if eyewitnesses adopt such a strategy, if answers could not be verified, it would thus *appear* as if recall were enhanced if *any* procedure were used that included instructions to try again. Importantly, none of these effects was differentially affected by the particular interview instructions given.

In this experiment the number of identifications made significantly correlated both before and after interview instructions with respective measures of both the number of correct identifications and the number of incorrect identifications, and thus negatively correlated with initial accuracy rate. Again this effect seemed to be independent of interview instructions. This latter effect was consistent with those found in Experiments 2 and 4 and to some extent with those of Experiment 8, though in this experiment the proportion of correct to incorrect responses was somewhat lower than in previous experiments, being 67/33 before interview instructions and lower still, 63/37, after interview instructions (interview conditions combined).

When hypnotic susceptibility was considered, overall, it was significantly correlated with the number of initial correct identifications, but not initial incorrect identifications, made before interview instructions, and between both the number of initial number of correct and incorrect identifications after interview instructions. Perhaps most interesting is the fact that this overall effect appeared to be more due to cognitive interview and control group than the hypnosis group. Thus, it would appear that hypnotisability has an effect even when (or even especially if!) hypnosis is not used. This finding may have been due to social desirability; i.e. subjects who wished to help the experimenter stated that they were more hypnotically susceptible and also paid more attention to the experimenter and the experimental instructions. Because of this increased attention compared to less hypnotically susceptible subjects they paid more attention to the stimuli, were able to initially identify more photographs and also expressed greater confidence. However, this tendency also encouraged them to recall more inaccurate responses after the interviewing instructions. Alternatively, or additionally, it is well established in the hypnosis literature that hypnotic susceptibility correlates significantly with measures of imaginative involvement (Spanos, 1986). It is possible that somehow a tendency to use visual imagery might help identification (it is perhaps relevant here that hypnotic susceptibility was not related to **recall**; where one might expect imagery skills to be less appropriate).

Further, more direct C-A relationships were investigated for total identifications. Broadly speaking, once again strong C-A relationships were found; this was entirely consistent with Experiment 8. Within-subjects C-A correlations were significant for each interview condition before and after interview instructions but no significant differences were found between interview conditions. In addition, the correlation of each subjects' average accuracy was significantly correlated before and after interview instructions in the control, cognitive interview and hypnosis conditions combined were significant before and after interview instructions. Also, correlations for the cognitive interview group before interview instructions and the control group after instructions were significant while the other correlations were also positive but not significant in the other cells (all *r*s $> .37$).

Between subjects C-A correlations before and after interview instructions were all not significant; this was consistent with Experiment 8. In addition, when average accuracy and average confidence score for each identification number was correlated very large correlations were produced. These correlations were significant for each interview condition before and after interview instructions.

When confidence ratings in correct and incorrect identifications were considered with the ANOVA analysis, subjects expressed significantly greater confidence in correct identifications than incorrect identifications. No difference was found between interview conditions but there was a significant main effect of interview instructions that indicated that confidence in both correct and incorrect identifications was increased after interview instructions. In addition, there was a significant interaction between interview condition and before/after interview instructions. This was because the hypnosis group expressed significantly greater confidence than the control and cognitive interview group in both correct and incorrect identifications. As such it is a similar finding to that mentioned previously in this experiment for the verified unverified comparisons.

These results suggest the limitations of judging accuracy solely on the basis of C-A relationships. Although the earlier analyses showed that interview instructions lowered the criterion for report, this did not apparently affect the C-A correlations. This was because it increased confidence in **all** responses both correct and incorrect; thus spurious increases in confidence for incorrect identifications were, in terms of C-A relationships, compensated for by increases in confidence for correct identifications.

No effects of interview condition or before after interview instructions was found for identifications rated as being 'absolutely certain' identifications, although once again the accuracy rate was very high (85%).

18.4.2. Questionnaire recall task

When the recall task was considered, no significant differences were found between interview conditions in terms of correct answers (as the paradigm was forced choice the effects for incorrect answers 'mirror' those for correct answers). When item difficulty was considered, a significant effect was found for the number of correct answers; i.e. easy questions were more likely to be answered correctly than medium questions, which in turn were more likely to be answered correctly than hard questions. These results again supported the validity of the categorisation of item difficulty.

One interaction between item difficulty and interview condition was significant; the difference between the number of easy and medium questions answered correctly was greater in the control group than in the hypnosis group.

It is not at all clear why such this effect occurred, especially as no differences between interview conditions were apparent when easy and medium items were considered independently. Without replication perhaps this should be dismissed as a statistical fluke.

Next, attention was given to within-subjects C-A correlations. These again support the view that, when questions which vary in difficulty are used, and thereby the probabilities of producing 'absolutely sure' and 'pure guess' responses are maximised, confidence accuracy relationships are considerably higher than have been previously been reported in the published literature (e.g. Kassin et al., 1989; Perfect et al., 1993). These findings are in general, similar to those found using this questionnaire in Experiment 7 and once again suggest that subjects are more confident about their correct answers than their incorrect answers.

C-A relationships appeared to be higher for easy items than for medium items which in turn tended to be higher than hard items. However, this was not significant for the ANOVA comparison of within-subjects correlations but this may only have been because the number of subjects that could be used in this analysis was only eight. This effect also occurred in Experiments 6 and 7, and thus it would seem to be reliable. As mentioned previously, one possible explanation is that different kinds errors were produced by the different question types. For easy questions, the majority of subjects provided correct answers and displayed high levels of confidence in the accuracy of their answers. Therefore, what few errors did occur may have been due to factors such as lapses of attention, or momentary distraction at the encoding stage. Consequently, those subjects who missed some easy stimuli would have expressed low confidence and showed poor accuracy to questions that concerned this information, thus producing a high C-A correlation for easy questions. However, for more difficult questions, even subjects who paid close attention to the film were unable to answer many of the questions correctly, and the few accurate responses they did make, many have resulted from guessing. Such guessing would again lower the C-A relationship for difficult items.

The above C-A relationships did not appear to be greatly influenced by the different interview conditions in this experiment. There was only one significant effect. When within-subjects C-A relationships for easy questions were considered independently, the mean C-A correlation for subjects in the cognitive interview condition was significantly greater than that for subjects in the control group. It may be possible to explain this finding in terms of Deffenbacher's (1980) optimality hypothesis. The cognitive interview condition, by enhancing the optimality of retrieval conditions via the four mnemonic instructions may also enhance eyewitnesses' ability to calibrate the accuracy of their recall and therefore increase the relationship between their confidence and their accuracy.

However, if this were so, why was no C-A enhancement found for the hypnosis condition which also included the cognitive interview mnemonics? Perhaps this may be explained by the differences in the way in which the mnemonics are used. In the case of the hypnotic condition the mnemonic instructions were given immediately after a hypnotic induction procedure in which they were repeatedly instructed to feel 'sleepy', 'tired' and 'relaxed' and also to remain 'hypnotised'. Perhaps optimal C-A calibration benefits not only from mnemonics, but, at least in the case of 'neutral' (i.e. not emotionally arousing) stimuli, an alert frame of mind.

When subjects' average confidence was correlated with their average number of correct answers, similar findings were produced as for the within-subjects analysis. Overall, subjects who expressed higher confidence in their answers were more likely to be accurate than subjects who expressed lower confidence. Again there appeared to be a trend for C-A relationships to be higher for easy items than for medium items which in turn were higher than for hard. None of these effects, however, was influenced by interview conditions.

When between-subjects correlations were considered, only one correlation was significant, for easy questions in the control group. However, it was not possible to determine whether this was due to the control instructions producing an enhancement of C-A relationships compared to the cognitive interview and hypnosis conditions (which also showed high C-A relationships on this measure) as direct statistical comparisons were impossible because of the small number of questions that it was possible to use in these calculations (four in the control condition, five in the cognitive condition group and six in the hypnosis condition). It should be noted that these particular correlations were based on a restricted range of items that did not include responses to the most easy and the most difficult questions (they excluded those items on which all subjects were correct or incorrect, because it was not possible to calculate C-A correlations for these particular questions).

In addition, the average accuracy rate and average confidence score for each question were correlated between-subjects. Very high, significant C-A correlations were found as for Experiment 7. This would seem to indicate that even within questions classified as easy, medium and hard there was still some variability in question difficulty and that average confidence scores for a question was related to the average number of correct responses on that question.

A positive relationship between subjects' confidence and their accuracy was also supported by the ANOVAs conducted on subjects' confidence in correct and incorrect answers. Consistently, confidence in correct answers was higher than confidence in incorrect answers, regardless of whether this was overall, or when easy, medium or hard items were considered independently. Once again no effects of interview condition were found.

Previously, it was suggested that in order to determine their confidence subjects may not only use an internal 'feeling of knowing' of whether their answer is correct to determine their confidence but also additional components. It was hypothesised that one such factor may be an individual's evaluation of how confident they *'should'* be. This hypothesis received some support. With respect to correct responses, higher confidence scores were reported for easy questions than for medium questions than for hard questions. However, with respect to confidence in incorrect answers the pattern was not so clear. Significantly higher confidence scores were given for medium questions answered incorrectly than for hard questions that were answered incorrectly. However, no other differences by question difficulty with respect to confidence in incorrect answers was found. No effects for interview conditions were found regarding the above.

When 'absolutely sure' responses were considered there was a again a very high accuracy rate; when a subject made an 'absolutely sure' response it was very likely to be correct. There were no differences because of interview condition in terms of number of correct and incorrect responses.

As noted previously, unlike in the identification task, hypnotic susceptibility did not appear to greatly influence confidence measures. One would have expected that if the results regarding hypnotic susceptibility and confidence in the identification task were due solely to demand characteristics, they would occur also on the recall task. However, this result may reflect differences in the cognitive effort required for recognition and recall tasks. Perhaps, the more active cognitive processes required for a recall task may make distract subjects from the social demands of the situation and so make them less susceptible to bias.

Taken as a whole, therefore, the results suggest that high C-A relationships are most likely to occur when a) the items to be remembered are relatively heterogeneous in terms of difficulty; b) the calculations are performed on aggregate scores, and c) subjects are 'absolutely sure' of their responses. Different interview instructions did not, in this experiment, dramatically affect C-A relationships, though there was some evidence that favoured the cognitive interview. However, it is plausible that some interview techniques may possibly have a greater influence on C-A relationships in more naturalistic settings (e.g. when contexts are changed).

18.4.3. Comparisons between recall questionnaire and identification questionnaire

The correlational analysis conducted between the recall questionnaire and the identification questionnaire showed little relationship between recall and identification performance. Subjects' C-A relationships, number of correct responses, and confidence in incorrect responses on the questionnaire task were not significantly related to the same measures on the identification task. This would seem to indicate that subjects' ability to perform on a recall task may not be strongly related to their ability to perform on a identification task and thus is consistent with the work of Wells in this area (Wells, 1985; Wells and Elizabeth, 1990). However, it must be noted that in these experiments almost all subjects were showing high C-A relationships on both tasks.

PART 3

CONCLUSIONS

OVERVIEW

Part three presents the conclusions drawn from the experimental work and gives suggestions for future work. It is divided into three parts.

Chapter 19 presents overall conclusions with regard to the verification programme.

Chapter 20 presents overall conclusions with regard to confidence-accuracy relationships.

Chapter 21 presents a critique of the thesis and describes the implications of the experimental programme for future work.

CHAPTER 19

CONCLUSIONS WITH RESPECT TO EXPERIMENTS IN THE VERIFICATION PROGRAMME

Conclusions regarding the experiments which included a verification factor are discussed here, including consideration of before/after instructions effects and the preliminary confidence measures. A more detailed account of confidence-accuracy relationships is given in the next chapter.

19.1 Verified/unverified comparisons

A number of hypotheses regarding verification were made, in particular: 1) subjects' whose answers could not be verified would estimate that they could accurately identify significantly more photographs than subjects whose answers could be verified; 2) subjects whose answers could not be verified would express significantly greater confidence in the accuracy of their estimates of how many photographs that they could accurately identify, and where applicable the accuracy of their actual identifications, than subjects whose answers could be verified; and 3) subjects' whose answers could not be verified would state that they would be prepared to identify more photographs with 'absolute certainty' in court than subjects whose answers could be verified.

Broadly speaking, despite motivating instructions, explicit information that answers could not be verified and powerful interview techniques such as hypnosis, no significant differences were found between verified and unverified groups in the hypothesised directions. However, there was one exception; for Experiment 2 subjects in the unverified group expressed significantly greater confidence in their estimates of how many photographs they could accurately identify than the verified group. However, given that this result was not replicated in any other experiment it would seem unwise to place much emphasis on this finding. Indeed, there were findings in the opposite direction to those hypothesised. In Experiment 5, where subjects were explicitly told that their answers could not be verified, the estimate of the number of photographs that subjects estimated that they could accurately identify was significantly *less* in the unverified group than the verified group. Also, no significant differences between interview conditions were found with regards to the verification variable.

Thus, taken as a whole, this evidence would seem to suggest that eyewitnesses may be reluctant to make up information simply because they are aware that answers that they provide cannot be checked. It is plausible that this finding may generalise to real-life situations; eyewitnesses may often be particularly reluctant to provide information that they believe might convict an innocent individual. This idea would seem to be borne out by the consistent finding across these experiments that subjects were very conservative in the number of identifications that they stated they could identify in court compared to the number of accurate identifications they estimated they could identify and the number of correct identifications they made.

However, it may also be the case that in real situations knowledge that answers cannot be verified may distort eyewitness performance through factors which are unrelated to the way in which an interview is conducted, and thus were not tested in the designs used here. For example, the eyewitness may be afraid of that he/she may find himself/herself the victim of reprisals from the criminal or the criminal's friends if he/she gives accurate testimony. Alternatively, the witness may have a good memory for the event, but may be reluctant to inform the police because, they will be required to give statements and possibly have to go to court which may require considerable time and effort on their part. In addition, witnesses who are aware that their testimony cannot be checked may alter their testimony to ensure that friends are not implicated in criminal proceedings. Such instances may be particularly difficult to detect as such individuals are unlikely to reveal themselves as friends of the accused. Thus, eyewitnesses are likely to be influenced not only by the interview procedure itself but also by wider considerations.

It could also be argued that the social demands in the present studies were less than might exist in a face to face police interview. However, there were no indications at all that increasing pressures to report produced a verification effect. Moreover, the present results do suggest that, other things being equal, memory motivation and facilitation instructions do not differentially affect reports according to verifiability.

19.2 Before/after instruction and preliminary confidence effects

When subjects' estimates of their performance, and their actual performance were considered, some significant effects of motivating instructions and interview techniques were found. The number of identifications that subjects estimated that they could make increased after motivating instructions or interview techniques. However, as noted previously, it is likely that this effect was at least in part due to repeated testing (see Nogrady, McConkey and Perry, 1985). In addition, average confidence in identifications was also increased after interview instructions when measures were taken, this was for control, cognitive interview and hypnosis conditions although subjects in the hypnosis condition slightly increased their confidence more than the other conditions. This finding for hypnosis is consistent with what might have been expected on the basis of the literature (see, for example, Wagstaff, 1989, 1993, 1995), because hypnosis, may possess characteristics that set it apart from other interview procedures (e.g. extra expectations for increasing accurate report). It also suggests that, unlike hypnosis, a cognitive interview procedure is no more susceptible to before/after confidence effects than a control for repeated testing.

Because before-after identifications and a forced-choice measures were taken in Experiment 9 it was possible to determine whether the above effects were due to real enhancements in recall or simply reflected a lowered criterion for report. As noted in the previous chapter, of particular interest, was the finding that both the number of (initial) correct and (initial) incorrect identifications were increased by interview instructions. Nevertheless, the ANOVAs for the forced-choice (total) correct identifications revealed no before/after effects whatsoever, regardless of interview instructions. Taken as a whole, this would suggest that if subjects are asked to repeat a memory task, they will lower their criterion for report. However, very importantly, lowering the criterion for report does not influence the accuracy of the reports in terms of the proportion of correct to incorrect answers. In other words the increase in information is not at the price of accuracy. Importantly, however, none of these accuracy effects was differentially affected by the particular interview instructions given. Nevertheless, these results do confirm that the extra increases in confidence found with hypnosis were possibly spurious.

It was also hypothesised that within the verified group there would be a positive relationship between subjects' estimates of how many photographs they thought they could accurately identify and the number of correct identifications they made. Broadly speaking this was supported as in all experiments by comparisons between subjects' estimates of how many accurate identifications they could make before and after interview instructions and the actual number of correct identifications that subjects made. In addition, across experiments subjects' estimates of how many photographs they could accurately identify was generally correlated with the number of correct identifications that they made. This appeared to be independent of interview conditions.

19.3 Hypnotic Susceptibility

In Experiment 8 hypnotic susceptibility correlated positively with the number of correct, but not incorrect identifications. And when it was considered in detail in Experiment 9, it was significantly correlated with the number of initial correct identifications, but not initial incorrect identifications, made before interview instructions, and between both the number of initial number of correct and incorrect identifications after interview instructions. Perhaps most interesting was the fact that this overall effect appeared to be more evident in the cognitive interview and control groups than the hypnosis group. Thus, it would appear that hypnotisability has an effect even when hypnosis is not used, and on the whole the effect would seem to be positive; i.e. it increases identification volume without necessarily showing decrements in terms of accuracy. As suggested previously this finding may have been due to social desirability; i.e. subjects who wished to help the experimenter may have stated that they were more hypnotically susceptible and also paid more attention to the experimenter and the experimental instructions. Because of this increased attention compared to less hypnotically susceptible subjects they paid more attention to the stimuli, were able to initially identify more photographs and also expressed greater confidence. However, in Experiment 9 this tendency also encouraged them to recall more inaccurate responses after the interviewing instructions. Alternatively, or additionally, as previously mentioned, it is well established in the hypnosis literature that hypnotic susceptibility correlates significantly with measures of imaginative involvement (Spanos, 1986). It is possible that somehow a tendency to use visual imagery might help identification. It is perhaps relevant here that hypnotic susceptibility was not related to recall; where one might expect imagery skills to be less appropriate.

19.4 Summary

To summarize so far, it seems that in terms of the present findings, verifiability did not adversely affect eyewitness reports, the motivating or memory facilitation instructions (including forcing subjects to guess) tended to increase the amount of both correct and incorrect information, but not necessarily at the price of accuracy in terms of the proportion of correct to incorrect responses. Interviewing techniques, of the kind used here, had little effect, except, relative to the others, hypnosis tended to slightly increase confidence without a corresponding increase in accuracy. Preliminary confidence-accuracy measures tended to show that most subjects were able to make fairly accurate estimates of their performance. And, finally, results indicated that hypnotic susceptibility is a possible correlate of eyewitness identification accuracy that may warrant further investigation.

If at all generalisable, perhaps these results suggest that, if the police want more information, if they are prepared to risk some increase in incorrect information, virtually **any** procedure which encourages witnesses to produce more information will be successful. However, there seems to little point in including hypnosis amongst these techniques, particularly when one considers the legal difficulties that have ensued from its application (Wagstaff, 1993).

CHAPTER 20

CONCLUSIONS WITH RESPECT TO CONFIDENCE ACCURACY RELATIONSHIPS

The experiments presented in this thesis consistently indicated a strong relationship between subjects' confidence in information which they provide and their accuracy. This relationship was particularly strong when within-subjects analyses were conducted on recall answers and item difficulty was varied. This would seem to indicate that eyewitnesses' confidence in responses that they provide is a good predictor of the likelihood of their responses being accurate. This contradicts many previous investigations of C-A relationships. However, the generalisability of these findings to real eyewitness situations assumes that experimental tasks described in this thesis are in some way comparable with eyewitness situations in real-life.

It could be argued that the item-difficulty paradigm that was used for the recall questionnaire that was presented here (i.e. asking questions of heterogeneous difficulty) cannot be generalised to real situations because in real situations questions are not used that vary in difficulty (i.e., they are homogenous). Such an argument would seem to highlight the fact that there is no objective measure of whether questions are likely to yield information that is useful to an investigation (i.e. forensically useful information). However, on the contrary, the use of questions that vary in difficulty may be a very appropriate way to investigate eyewitness C-A relationships. The utility of information in 'solving' crimes, or 'forensic utility', is not necessarily proportionally related to item difficulty. The idea that questions must be difficult in order to obtain useful information may be misleading. An example that was given earlier concerned establishing the gender of a suspect. This may be relatively easy, yet it is of considerable use to an investigation, immediately eliminating half the population from future enquiries. Thus, information gathered from easy questions may be very relevant to eyewitnesses' testimony.

Newlands and George (1994) found that police officers were consistent in their judgements of what are 'good' and what are 'poor' witness statements. It is also plausible that police officers have consistent ideas of what information has forensic utility. Future work may seek to determine 1) if such target forensic information exists, and 2) if so, does a strong C-A relationship remains when forensic utility is considered.

In addition, item difficulty may affect whether or not misleading post-event information or leading questions are incorporated into eyewitness reports. For example, one would assume that eyewitnesses are very resistant to leading questions that concern easy items as they are very confident that they knew the correct answer and are aware that this is different from that suggested by the misleading post-event information or the leading question. For difficult questions such decisions may be more difficult, subjects may be more susceptible to errors concerning misleading post event information or leading questions as they are not so sure of alternative responses.

Confidence-accuracy relationships were also calculated for the identification task. Again strong C-A relationships were found when within-subjects correlations were considered, although not when between-subjects correlations were considered. These findings may be less applicable to forensic situations than those produced with the recall questionnaire. For example, in real eyewitness situations one rarely has to remember 20 or more target faces. In addition, in more conventional experiments C-A relationships are calculated with line-up presentations, in which only one target individual is presented or the target individual is absent. It can be noted that the majority of research on C-A relationships that have used these paradigms and found little C-A relationship (see, for example, Bothwell, Deffenbacher, & Brigham 1987; Wells & Murray, 1984; but see Sporer et al., in press), so it is possible that the C-A relationship for single identifications may be weaker than for the kinds of items used in the recall questionnaire. On the other hand, it is also possible that the same principles of item difficulty may govern both situations; perhaps C-A relationships for identifications may be increased if identification difficulty is varied, by, for instance, varying exposure time, illumination, and time before the identification task.

The correlational analysis conducted between the recall questionnaire and the identification questionnaire showed little relationship between recall and identification performance. Thus, it would appear that subjects' recall may not necessarily be an appropriate indicator of the likelihood of them making an accurate identification. However, it should be noted that in this experiment the target information for the recall questionnaire concerned different stimuli to that of the identification questionnaire. It is therefore possible that if the recall task and the identification task concerned the same target (e.g. the same individual), then the two factors may be more closely related; although previous studies would seem to suggest that this is not so (Wells, 1985; Wells and Elizabeth, 1990).

In many respects the C-A relationships presented in this thesis may represent the best C-A relationships which are likely to be found. Other variables in 'real-life' situations may reduce these relationships. For example, in 'real-life' situations eyewitnesses are not required to rate their own confidence on a Likert scale, instead their confidence is gauged by investigators or jurors through a variety of other factors such as tone of voice, eye contact as well as requesting eyewitnesses to express how sure they are of an answer. Further factors, such as greater social pressure in interview or trial situations may make eyewitnesses less or more confident in parts of their statements (dependant on what kind of pressure is applied) and longer retention intervals may change the C-A relationship.

As noted previously, memory facilitation procedures had little overall effects on C-A relationships. However, it was noted that in more realistic conditions perhaps the cognitive interview, by enhancing retrieval cues in conditions where there is a greater discrepancy between encoding and retrieval conditions, may show advantages in terms of increased C-A relationships over alternative procedures.

Of interest, was the fact that hypnosis did not appear to have a great adverse influence on subjects' C-A relationships on the recall questionnaire. This would seem to support a growing literature which suggests that hypnosis does not necessarily produce adverse effects (Wagstaff, 1993, 1995). However, as previously mentioned, there was a subtle effect of hypnosis on the average confidence which subjects expressed in the total number of identifications that they made. There was a significant before after effect of hypnosis such that confidence was increased in both correct and incorrect identifications compared to the cognitive interview and a control group. However, this result should not be overemphasised as it must be borne in mind that there was no significant difference between interview conditions in terms of confidence in identifications both before or after interview procedures.

The relationship between confidence and accuracy found in this thesis may have implications for the ways in which police investigations are conducted and jurors' decisions are made. As such, it may be appropriate to evaluate these findings in the terms that were outlined in Chapter 1 by Wells (1978). He suggested that the role of eyewitness testimony research is:

> ... to generate scientific knowledge that will maximize the chances that a guilty defendant will be justly convicted while minimizing the chances that an innocent defendant will be mistakenly convicted. (Wells, p.1546)

It has been suggested here, that C-A relationships may be much higher than has previously reported. If C-A relationships are exaggerated, this may lead the police to over rely on eyewitness confidence to predict accuracy (if that is not already being done). As a result, innocent individuals may be convicted of crimes that they did not commit because of a confident but inaccurate witness. However, it is also likely that the current consensus that there is little relationship between eyewitnesses confidence and their accuracy, if communicated to the police and courts (e.g. Canter, 1994; Kassin, Smith and

Ellsworth, 1989), might lead to individuals *not* being convicted of crimes that they did commit because investigating officers or jurors are told not to use confidence to predict accuracy.

Clearly, neither high levels of false convictions nor large numbers of criminals escaping justice and refunding are desirable. Thus, although gross overreliance on eyewitnesses' confidence in order to predict accuracy may lead to the inappropriate conviction of innocent individuals, ignoring eyewitness confidence as a predictor of accuracy may mean that guilty individuals are not convicted of crimes that they did commit. It would seem appropriate to draw the tentative conclusion that reliance on an eyewitness's confidence may be more useful in determining his/her accuracy than psychologists have recently assumed.

CHAPTER 21

CRITIQUE OF THESIS AND FUTURE WORK

The following chapter concerns suggestions for appropriate future work. Many of these suggestions could have been incorporated within the present thesis, thus, this chapter also includes a number of criticisms of this thesis.

To reiterate, the main findings of this thesis are as follows.

1) There is, on the whole, a strong relationship between eyewitnesses' confidence and their accuracy.

2) Verifiability does not adversely affect eyewitness reports.

3) Motivating or memory facilitation instructions (including forcing subjects to guess) tend to increase the amount of both correct and incorrect information, but not necessarily at the price of accuracy in terms of the proportion of correct to incorrect responses.

4) Interviewing techniques, of the kind used here, have little effect on any of the measures, except hypnosis tends to slightly increase confidence without a corresponding increase in accuracy.

5) Hypnotic susceptibility is a possible correlate of eyewitness identification accuracy that may warrant further investigation.

However, there are obvious limitations to the kind of studies presented here in terms of how generalisable they are to real-world interview situations. Unfortunately, of course it is impossible to find parallels between the designs used here and interviews in the real world, and it would be impossible to investigate the interactions between variables examined in real-life situations (e.g. archival material is not going to reveal simultaneous comparisons between hypnosis, the cognitive interview and control conditions etc.). Nevertheless, there are a number of ways in which this thesis could have been directed, and future work should be directed, to make results more generalisable.

21.1 Forensic relevance

The purpose of a police interview is not necessarily to produce as much information as possible from the witness; it has a number of specific aims and objectives. These are; 1) to discover if a crime has been committed and if so which crime; 2) if a crime has been committed to find evidence to identify the individual responsible; 3) to produce evidence that prevents a *guilty* criminal using an inappropriate defence, and 4), to determine if the eyewitness is not telling the truth.

Initially an interviewing officers must determine if a crime has actually has occurred, for example, a complainant may report an incident to the police that is not actually a criminal act. However, once it has been established that a criminal act has occurred police officers must decide which laws have been broken and collect evidence that shows that this offence has taken place. To determine what offence has taken place the investigating officers may also have to consider the intent of a criminal. For example, if an individual is killed on the spur of the moment the offence is manslaughter, while if the killer planned the offence, the offence becomes one of murder.

Once it is apparent that a criminal act has taken place police officers have to identify the criminal responsible. This may be achieved by obtaining descriptive information from eyewitnesses, or if a suspect has been apprehended, asking the witness to attend an identification parade. Other sources of information may also be used to capture a suspect, these are outlined later.

Officers conducting an investigation must also identify inappropriate lines of defence that a criminal may take. For example, in a manslaughter case a criminal may justify his/her actions by claiming that he/she acted in self-defence. Thus, investigating officers should ask witnesses if the suspect was defending himself/herself. If they suggest that this was not the case it may prevent or undermine such a claim later in court. Of course if the witnesses suggest that the killing was in self-defence this may influence what charges, if any, are made. Similarly, the investigating officer must try to prevent the defence of 'It wasn't me', by, for example, ensuring that appropriate procedures are used throughout the investigation (e.g. fair identification parades in accordance with the Codes of Practice). The investigating officers must also try to ascertain if the eyewitness is lying (e.g. to try to prevent a friend getting into trouble, or because he/she is afraid of retribution if he/she tells the truth).

In addition, as outlined in the introductorary chapters, reliance on eyewitness testimony will vary from case to case depending on what other information the police can use (i.e. suspect confession, fingerprints).

Thus, to effectively examine the practical effects of an item of interest (e.g. C-A relationships, effects of verification, effects of interview instructions) it is essential to relate this to the quality of information that is produced that is consistent with these aims. While attempts have been made to class information according to how useful or critical it would be to an investigation (for example, Geiselman, et al., 1985), what an experimenter codes in an arbitrary way as central information is not necessarily what the police themselves might find most useful for the successful completion of criminal investigations. Any empirical evaluation of eyewitness testimony should therefore take into account forensic relevance.

Preliminary work by Newlands and George (1994) appears to indicate that police officers are consistent in their perceptions of what is a good and what is a bad witness statement. Thus, it could have been possible in this thesis, and in future work it may be possible, to use police officers' evaluations to assess eyewitness performance in a way that is more applicable to real-life settings.

21.2 Verification

Little effect was shown for verification on eyewitnesses' performance in the present thesis. Indeed, with hindsight, too much emphasis was given to the area of verification in this thesis. Given the lack of significant differences between verified and unverified groups it would have been more productive to have stopped the verification programme at an earlier stage. Instead, efforts could have been switched to more productive research areas as are outlined in this chapter.

However, with regards to verification, as stated previously, in real life situations, factors outside of the demands of the interview may adversely influence eyewitnesses' performance when verification is not possible (e.g. an eyewitness may be afraid of that he/she may find himself/herself the victim of adverse repercussions from the criminal or the criminal's friends if he/she gives accurate testimony). If so, then it might be the case that the real effects of verification are to a large extent beyond the scrutiny of experimental research. It might be possible, however, to examine the unverified/verified distinction through archival material using a technique such as Statement Reality Analysis (Gudjonsson, 1992).

21.3 Confidence-Accuracy relationships

As noted previously, the present thesis could have paid (and future evaluations of C-A relationships should pay) attention to the forensic usefulness of information that is provided to see if the high C-A relationships that have been reported here remain with forensically relevant criteria.

In addition, it may be appropriate to evaluate subjects' C-A relationships in terms of others perceptions' of eyewitnesses' accuracy in given statements. It appears that in the majority of investigations and trials, what influences investigating officers and jurors is not so much how confident an eyewitness *is*, but how confident he/she is *perceived* to be. A future experiment may wish to compare observers ratings' of how confident they perceive subjects to be in relation to their accuracy.

21.4 Interview techniques

The most obvious criticism of the present research is that what the writer has called the 'interview' conditions were not true interviews at all; i.e. in real life interviews are conducted face to face and involve an interchange of questions and answers between interviewer and interviewee. This is a valid criticism and perhaps greater steps could have been taken to include more naturalistic, interpersonal factors. However, there are also problems associated with such stratergies.

One of the fundamental problems with evaluating the cognitive interview in its most recent forms is that it is no longer strictly a 'procedure', rather it is a somewhat amorphous collection of techniques. Currently a number of interview techniques are described as being cognitive interview techniques. For example, the 'cognitive interview' that Merseyside Police are trained in is the technique described by Fisher and Geisielman (Fisher, Geiselman & Amador, 1989), and the technique that the Thames Valley Police are trained in is also called the cognitive interview although the change perspectives mnemonic is not used. However, the cognitive interview that Fisher is currently describing as the cognitive interview emphasises the social processes to a greater extent then retrieval mnemonics (Fisher, personal communication), and thus he describes what Memon et al. term a 'structured interview' a 'cognitive interview'. George's (1990) police officers who described themselves as using cognitive interviews only regularly used the reinstate context and report all mnemonic, while the cognitive interview that was used in these experiments that remains faithful to the original (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian and Prosk, 1984) cognitive interview. And finally, there is forensic hypnosis; typically this is very similar to the enhanced cognitive interview (Fisher, Geiselman & Amador, 1989; Wagstaff, 1992). Does this mean that hypnosis it is, in fact, a cognitive interview too? It would now seem that **any** interview that attempts to overcome what are seen to be deficiencies in 'standard' interviewing procedures could be termed 'cognitive'.

Thus, it would no longer seem to be the case that any one technique may claim to be a cognitive interview. Consequently, it may be more appropriate to describe what the technique actually consists of, so that the reader is able to appropriately evaluate the results.

To some extent this highlights a general problem with research on the cognitive interview. Most research has looked at the cognitive interview as a 'package'; but very little attempt has been to evaluate the efficacy of the individual components (Bekerian and Dennett, 1993). We can now see a justification for the approach adopted in the present thesis. Given the huge range of variables that could be investigated in comparing interview techniques (including effects such as rapport, the number of interruptions, the kinds of questions asked etc., etc.), it would have been impossible to have systematically examined all of these at once. Instead, the experiments concentrated on the specific factor of the specific instructions given to motivate witnesses to remember or facilitate their memory, whilst holding the vast range of other social variables constant.

The implication of this approach for future research is obvious. A **componential** analysis of interview techniques is needed to judge what is effective and what is not. For instance, recently Boon and Noon (1994) looked at the individual effectiveness of the four mnemonic components, report all, change order, change perspective, and reinstate context, compared with a simple control 'try again' instruction. All were effective in increasing accurate recall except the change perspective mnemonic, and none increased inaccurate recall. However, no systematic attempt was made in this study to investigate the efficacy of various combinations of these mnemonics, to see whether the effects are additive. If they are not additive, then just one or two might suffice. Certain combinations might even act against each other. Moreover, no attention was paid in this study to the relative or combined contribution of other elements in the 'enhanced cognitive interview', in particular, increasing rapport, using witness compatible questioning, and improving motivation. Thus further, componential analysis would seem particularly important. Moreover, a comprehensive componential evaluation of interview techniques could potentially offer substantial savings in terms of length and complexity of training of interviewers, and time spent interviewing witnesses; problems that have so far limited the extensive application of the cognitive interview.

At the same time, the issues of practical application and forensic relevance should not be ignored. For instance, the problems with conventional interviewing techniques may not actually be as clear cut as they at first seem.

Informal discussions with Police officers suggest that there are specific reasons for officers to use interruptions and a question/answer format (c.f. chapter 3). Some Police officers say that they will use such strategies deliberately to curtail an eyewitness's report in certain situations (i.e. to limit the eyewitness's report only to that which the officer deems necessary). Such situations, might include when an officer has to interview a number of eyewitnesses. If the officer is at a crime scene and is confronted with a number of witnesses who require interviewing, he or she must interview each quickly to ensure that all are interviewed. If a long time is spent interviewing each witness the officer risks other potentially valuable witnesses leaving the crime scene. Thus, a method of conducting a rapid, succinct, interview can sometimes be more useful than a technique which produces a lengthy, more complete account. A similar situation, where a brief interview is necessary is when a Police officer has already has been given another crime to deal with after a current one and so has to rapidly respond to that request. Thus, the above factors evident in many Police interview procedures may not always be as disadvantageous as Fisher et al. (1987) have suggested; it may depend on the situation.

When considering Police interview techniques it should also be noted that Police officers do not just interview victims and bystanders, they also spend a considerable time interviewing suspects. This may impact on the interviewing techniques which they use. For example, asking a closed question which requires a rapid answer may not produce the best testimony from an eyewitness because it does not encourage 'focused' retrieval and an elaborated response; but such a technique may lead a suspected criminal to make an incriminating reply precisely because he or she was not given the time to think of a false statement to give. Thus the requirements of a 'good' interview of a suspect may be considerably different to those of a 'good' interview of a non-hostile eyewitness. Some Police officers report that if they alternate from interviewing a suspect to interviewing an eyewitness in a short period of time it is difficult to switch quickly from the frame of mind necessary to conduct a 'good' interview of a suspect to that necessary to conduct a 'good' interview of an eyewitness. Thus, they find themselves interviewing eyewitnesses in an inappropriate manner. This may explain some of the problems identified by Fisher et al. Further, it may have practical implications for Police interviewing procedures; perhaps some Police officers should specialise in eyewitness interviewing while others specialize in the interviewing of suspects.

George (1990) noted that Police officers who had not been formally trained in interviewing techniques "mysteriously share a common schema for deriving information". He states:

> "In the absence of training... Police officers somehow all perform and acquire information in the same manner. Intuitively, as experienced practitioners, it was tempting previously to assume that police necessarily perform this task in the most effective manner possible" (p.125).

These standard techniques are similar to those outlined by Fisher et al. (1987). The fact that a 'common schema' exists may itself indicate something of importance. Perhaps, on the basis of experience, this schema has proved on balance to be very effective; that is to say, the fact that many Police officers use a similar style may indicate that standard Police techniques may have considerable utility. Given the practical considerations just identified, standard Police interviewing techniques may represent a good *compromise* interview method; i.e. a useful general framework with which to 1) interview suspects, 2) interview eyewitnesses (sometimes in a limited period of time) and 3) be able to take appropriate notes. Moreover, it may suit resourcing policies; there seems little point in gathering a large amount of information for a minor crime if there are not enough resources to follow up leads anyway. However, for certain *specific* situations it may not be the most appropriate method. For example, when a serious crime has been committed more resources, especially time, are usually allocated. In such instances it maybe more appropriate to use a technique which may take longer but generate more information, such as the cognitive interview.

Future research, therefore, might more usefully be guided by what the police need, rather than what psychologists **think** the police need.

In addition, future research might usefully address the importance of eyewitness' perceptions concerning how well the Police officer interviewed them. Indeed, the interview of an eyewitness provides a good opportunity to enhance Police-public relations (Fisher & Geiselman, 1992). A well conducted interview with a concerned and genuinely interested Police officer is likely to increase the eyewitness's respect for and future willingness to co-operate with the Police. The alternative, a poorly conducted interview with an unconcerned Police officer, showing little interest in what the eyewitness says, may discourage the eyewitness from future support of the Police. As a crime incident is likely to be a talking point for a long period of time, then the eyewitness's impression of the Police is liable to be relayed to a large number of people. Thus, it may have a large impact on a locality. Perhaps here then is an unanticipated benefit of the cognitive interview which may also produce a positive contribution to a forensic investigations.

REFERENCES

Azjen, I. (1988). Attitudes, Personality and Behaviour. Milton Keynes: Open University Press.

Adams-Price, C. (1992). Eyewitness memory and aging: Predictors of accuracy in recall and person recognition. Psychology and Aging, 7, 602-608.

Ainsworth, P.B., & King, E. (1988). Witnesses' perceptions of identification parades. In M.M. Gruneberg, P.E. Morris, R.N. Syke (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 67-70. Chichester: John Wiley & Sons.

Allen, V.L., & Wilder, D.A. (1980). Impact of group consensus and social support on stimulus meaning: Mediation of conformity by cognitive restructuring. Journal of Personality and Social Psychology, 39, 116-1124.

Allwood, C.M. (1994). Confidence in own and others' knowledge. Scandanavian Journal of Psychology, 35, 198-211.

Alper, A., Buckhout, B., Chern, S., Harwood, R., & Slomovits, M. (1976). Eyewitness identification: Accuracy of individual v's composite recollection of a crime. Bulletin of the Psychonomic Society, 8, 147-149.

Anderson, R.C., & Pichert, J.W. (1978). Recall of previously unrecallable information following a shift in perspective. Journal of Verbal Learning, 17, 1-12.

Anderton, C.H. (1986). The forensic use of hypnosis. In F.A. DePiano and H.C. Saltzberg (Eds), Clinical Applications of Hypnosis. Norwood, New Jersey: Ablex.

Asch, S. (1957). Effects of group pressure upon the modification and distortion of judgement. In H. Guetzkow (Ed), Groups, Leadership and Men. Pittsburgh, PA: Carnegie.

Aschermann, E., Mantwill, M., & Kohnken, G. (1991). An Independent Replication of the Effectiveness of the Cognitive Interview. Applied Cognitive Psychology, 5, 489-495.

Atkinson, R.P., & Hustead, R.J. (1992). Moderating influences of hypnotic susceptibility on signal-detection task performance. Current Psychology: Research and Reviews, 11, 195-202.

Ault, R.L. (1979). FBI guidelines for use of hypnosis. The Journal of Clinical & Experimental Hypnosis, 27, 449-451.

Baddeley, A. (1988). But what the hell is it for? In M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds), Practical Aspects of Memory: Current Research & Issues, pp. 3-18. Chichester: John Wiley & Sons.

Baker, R.A., Haynes, B., & Patrick, B.S. (1983). Hypnosis, memory and incidental memory. American Journal of Clinical Hypnosis, 25, 253-262.

Baranski, J.V., & Petrusic, W.M. (1994). The calibration and resolution of confidence in perceptual judgements. Perception and Psychometrics, 55, 412-428.

Barber, T.X. (1991). The locksmith model: Accessing hypnotic responsiveness. In S.J. Lynn & J.W. Rhue (Eds), Theories of Hypnosis: Current Models and Perspectives, pp. 253-262. New York, USA: Guildford..

Barber, T.X. (1965). Hypnotic age regression: A critical review. Psychosomatic Medicine, 24, 286-299.

Barber, T.X. (1969). Hypnosis: A scientific approach. New York: Pergamon.

Barber, T.X., & Calverley, D.S. (1966). Effects on recall of hypnotic induction, motivational suggestions, and suggested regression: A methodological and experimental analysis. Journal of Abnormal Psychology, 71, 169-180.

Barber, T.X., & Wilson, S.C. (1978). The Barber suggestibility scale and the creative imagination scale: Experimental and clinical applications. The American Journal of Clinical Hypnosis, 21, 84-108.

Barnier, A.J., & McConkey, K.M. (1992). Reports of real and false memories; The relevance of hypnosis, hypnotizability, and context of memory text. Journal of Abnormal Psychology, 101, 521-527.

Baumeister R.F. & Newman, L.S. (1994). Self-regulation of congnitive inference and decision processes. Personality and Social Psychology Bulletin, 20, 3-19.

Behrman, B.W., & Vayder, L.T. (1994). The biasing influence of a police showup: Does the observation of a single suspect taint later identification? Perceptual and Motor Skills, 79, 1239-1248.

Bekerian, D.A. (1993). In search of the typical eyewitness. American Psychologist, 48, 574-576.

Bekerian, D.A., & Dennett, J.L. (1988). Memory on trial. In M.M. Gruneberg, P.E. Morris & R. N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 95-99. Chichester: John Wiley & Sons.

Bekerian, D.A., & Dennett, J.L. (1993). The Cognitive Interview: Reviving the issues. Applied Cognitive Psychology, 7, 275-297.

Bekerian, D.A., Dennett, J.L., Reeder, C., Sloper, K., Saunders, H., & Evans, L. (1994). The influence of the Cognitive Interview technique on attributions of memory. Poster presented at the BPS Annual Conference, Brighton, UK.

Bell, B.E., & Loftus, E.F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. Journal of Personality and Social Psychology, 56, 669-679.

Belli, R.F., Lindsay, D.S., Gales, M.S., & McCarthy, T.T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. Memory & Cognition, 22, 40-54.

Benedek, E.P., & Schetky, D.H. (1987). Problems in validating allegations of sexual abuse. Part 1: Factors affecting perception and recall of events. Journal of American Acad. Child Adol. Psychiatry, 26, 912-915.

Berger, J.D., & Herringer, L.G. (1989). Individual differences in eyewitness recall accuracy. Journal of Social Psychology, 13, 807-813.

Berry, D.S., & Wero, J.L.F. (1993). Accuracy in face perception: A view from ecological psychology. Journal of Personality, 64, 497-519.

Blake, M. (1973). Prediction of recognition when recal fails: Exploring the feeling of knowledge phenomenon. Journal of Verbal Learning and Verbal Behaviour, 12, 311-319.

Blanchet, G., & Blanchet, A. (1994). Interactional effects of the environment on interviews. European Journal of Psychology of Education, 9, 41-53.

Boon, J., & Noon, E. (1994). Changing perspectives in cognitive interviewing. Psychology, Crime & Law, 1, 59-69.

Boon, J.C.W., & Davies, G. (1988). Attitudinal influences on witness memory: Fact and fiction. In M.M. Gruneberg, P.E. Morris, R.N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 53-58 Chichester: John Wiley & Sons.

Bothwell, R.C., Brigham, J.C., & Pigott, M.A. (1987). An exploratory study of personality differences in eyewitness memory. Journal of Social Behaviour and Personality, 2, 335-343.

Bothwell, R.C., Deffenbacher, K.A., & Brigham, J.C. (1987). Correlation of Eyewitness Accuracy and Confidence: Optimality Hypothesis Revisited. Journal of Applied Psychology, 72, 691-695.

Bowers, K.S. (1983). Hypnosis for the Seriously Curious. New York: Norton.

Bowers, K.S., & Davidson, (1991). A neodissociative critique of Spanos's social psychological model of hypnosis. In S.J. Lynn & J.W. Rhue (Eds). Theories of Hypnosis: Current Models and Perspectives. New York: Guildford.

Bowers, K.S., & Hilgard, E.R. (1988). Some complexities in understanding memory. In Pettinati HM (Eds), Hypnosis and Memory, pp. 3-18. New York: Guildford Press.

Brigham, J.C. (1988). Is witness confidence helpful in judging eyewitness accuracy? In M.M. Gruneberg, P.E. Morris, R.N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 77-82. Chichester: John Wiley & Sons.

Brigham, J.C., and Bothwell, R.K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. Law and Human Behaviour 19-30.

Brigham, J.C., Maass, A., Snyder, L.D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. Journal of Personality and Social Psychology, 42, 673-681.

Brigham, J.C., & Wolfskeil, M.P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitnes identifications. Law and Human Behaviour, 7, 337-349.

Brown, E., Deffenbacher, K., & Sturgil, W. (1977). Memory for faces and the circumstances of encounter. Journal of Applied Psychology, 62, 311-318.

Buckhout, R., & Regan, S. (1988). Explorations in research on the other-race effect in face recognition. In M.M. Gruneberg, Morris P.E., Sykes R.N. (Eds), Practical Aspects of Memory: Current Research and Issues. pp. 40-46 Chichester: John Wiley & Sons.

Buckhout, R. (1974). Eyewitness testimony. Scientific American, 321. 23-31.

Buckhout, R., Eugenio, P., Licitra, T., Oliver, L., & Kramer, T.H. (1981). Memory, hypnosis and evidence. Social Action and the Law, 7. 67-72.

Bull, D., & Stratta, E. (1994). Police community consultation: An examination of its practice in selected constablaries in England and New South Wales Austrailia. Austrailian and New Zealand Journal of Criminology, 27. 237-249.

Bull, R. (1978). Eyewitnesses also have ears. In M.M. Gruneberg, P.E. Morris, Sykes R.N. (Eds), Practical Aspects of Memory. pp. 211-218. London: Academic Press.

Bull, R. (1992). Obtaining evidence expertly: The reliability of interviews with child witnesses. Expert Evidence, 1. 5-12.

Bull, R. (1993). Innovative techniques for the questioning of child witnesses especially those who are young and those with learning disability. Paper presented to the Annual Kent Psychology Forum, Ohio, USA.

Bull, R., & Clifford, B. (1979). Eyewitness memory. In M.M. Gruneberg & P.E. Morris (Eds), Applied Problems in Memory. pp. 151-184.London:Academic Press.

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering Emotional Events. Memory and Cognition, 20. 277-290.

Busemeyer, J.R. (1993). Violations of the speed-accuracy tradeoff relation: Decreases in decision accuracy with increases in decision time. In O. Stevenson & A. J. Maule (Eds), Time Pressure and Stress in Human Judgement and Decision Making. pp. 181-193. New York: Plenum Press.

Canter, D.V. (1993). The wholistic, organic researcher: Central issues in clinical research methodology. In G. Powell, R. Young & S. Frost (Eds), Curriculum in Clinical Psychology. pp. 40-56. Leicester: BPS.

Canter, D.V. (1994) Criminal Shadows Swindon: Soft Back Preview.

Ceci, A.J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. Psychological Bulletin, 113, 403-439.

Ceci, S.J., Loftus, E.F., Leichtman, M.D., & Bruck, M. (1994). The possible role of source misattributions in the creation of false beliefs among preschoolers. The International Journal of Clinical and Experimental Hypnosis, 42, 304-320.

Chandler, C.C. (1994). Studying related pictures can reduce accuracy, but increases confidence, in a modified recognition test. Memory and Cognition, 22, 273-280.

Chaves, J.F. (1989). Hypnotic control of clinical pain. In N.P. Spanos & J.F. Chaves (Eds), Hypnosis: The Cognitive Behavioural Perspective. Buffalo, New York: Prometheus.

Chaves, J.F. (1992). Role-theoretical and cognitive-behavioural perspectives on hypnosis. Contemporary Hypnosis, 9, 16-20.

Christianson, S.A. (1992). Emotional Stress and Eyewitness Memory: A Critical Review. Psychological Bulletin, 112, 284-309.

Christianson, S.A., Loftus, E.F., Hoffman, H., & Loftus, G.R. (1991). Eye Fixations and Memory for Emotional Events. Journal of Experimental Psychology: Learning, Memory and Cognition, 17, 693-701.

Christianson, S.A., & Hubinette, B. (1993). Hands up! A study of witnesses' emotional reactions and memories associated with bank robberies. Applied Cognitive Psychology, 7, 365-379.

Classen, C., Koopman, C., & Spiegel, D. (1993). Trauma and dissociation. Bulletin of the Menninger Clinic, 57, 178-194.

Clifford, B. (1978). A critique of eyewitness research. In M.M. Gruneberg, P.E. Morris, R.N. Sykes (Eds), Practical Aspects of Memory, pp. 199-208 London: Academic Press.

Clifford, B.R., & Hollin, C.R. (1981). Effects of the Type of Incident and the Number of Perpetrators on Eyewitness Memory. Journal of Applied Psychology, 66, 364-370.

Clifford, B.R., & Richards, V.J. (1977). Comparison of recall of policemen and civilians under conditions of long and short durations of exposure. Perceptual and Motor Skills, 45, 503-512.

Clifford, B.R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. Journal of Applied Psychology, 63, 352-359.

Coe, W.C., & Sarbin, T.R. (1991). Role theory: Hypnosis from a dramaturgical and narrational perspective. In S.J. Lynn & J.W. Rhue (Eds), Theories of Hypnosis: Current Models and Perspectives. New York: Guilford.

Cohen, G., & Faulkner, D. (1989). Age differences in source forgetting: effects on reality monitoring and on eyewitness testimony Psychology and Aging, 4, 10-17.

Cooper, L.M., & London, P. (1973). Reactivation of memory by hypnosis and suggestion. International Journal of Clinical and Experimental Hypnosis, 29, 312-323.

Crawford, H.J., & Allen, S.N. (1983). Enhanced visual memory during hypnosis as mediated by hypnotic responsiveness and cognitive stratergies. Journal of Experimental Psychology: General, 112, 662-685.

Crawford, H.J., Brown, A.M., & Moon, C. (1993). Sustained attentional and disattentional abilities: Differences between low and high hypnotizable persons Journal of Abnormal Psychology, 102, 534-543.

Cronin, O., Eaves, R., Kupper, B., Memon, A., & Bull, R. (1992). The cognitiive interview: A child centred approach? Paper presented at the NATO Advanced Study Institute: The Child Witness in Context: Cognitive, Social and Legal Perspectives, Lucca, Italy.

Cronin, O., Memon, A., Eaves, R., Kupper, B., and Bull, R. (1992). The cognitive interview with child witnesses: A child centred approach?. Paper presented at the Third European Conference of Psychology and Law, Oxford, UK.

Cutler, B.L., & Penrod, S.D. (1988a). Juror decision making in eyewitness identification cases. Law and Human Behaviour, 12, 41-55.

Cutler, B.L., & Penrod, S.D. (1988b). Context reinstatement and eyewitness identification. In G.M. Davies & D.M. Thomson (Eds), Memory in Context: Context in Memory, pp. 231-24.4 Chichester: John Wiley & Sons.

Cutler, B.L., & Penrod, S.D. (1989). Forensically relevant moderators of the relation between eyewitness identification accuracy and confidence. Journal of Applied Psycholgy, 74, 650-652.

Cutler, B.L., Penrod, S.D., & Dexter, H.R. (1990). Juror sensitivity to eyewitness identification evidence. Law and Human Behaviour, 14, 185-191.

Cutler, B.L., Penrod, S.D., O'Rourke, T.E., & Martens, T.K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. Social Behaviour, 1, 113-134.

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987). Improving the reliability of eyewitness identification: Putting context into context. Journal of Applied Psychology, 72, 629-637.

Cutler, B.L., Penrod, S.D., & Stuve, T.E. (1988). Juror decision making in eyewitness identification cases. Law and Human Behaviour, 12, 41-55.

Das, J.P. (1961). Learning and recall under hypnosis and in the wake state: A comparison. Archives of General Psychiatry, 4, 517-521.

Davies, G., Tarrant, A., & Flin, R. (1989). Close Encounters of the witness kind: Children's memory fo simulated health inspection. British Journal of Psychology, 80, 415-429.

Davies, G.M., Ellis, H.D., & Shepherd, J.W. (1978). Face recognition as a function of mode of representation. Journal of Applied Psychology, 63, 180-187.

DePiano, F.A., & Salzberg, H.C. (1981). Hypnosis as an aid to recall of meaningful information presented under three types of arousal. International Journal of Clinical and Experimental Hypnosis, 29, 383-400.

Deffenbacher, K.A. (1980). Eyewitness accuracy and confidence. Law and Human Behaviour, 4, 243-260.

Deffenbacher, K.A. (1988). Eyewitness research: The next ten years. In M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 21-26. Chichester: Wiley.

Deffenbacher, K.A. (1991). A maturing of research on the behaviour of eyewitnesses. Applied Cognitive Psychology, 5, 377-402.

Deffenbacher, K.A., Brown, E.L., & Sturgill, W. (1978). Some predictors of eyewitness memory accuracy. In M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds), Practical Aspects of Memory, pp. 219-226. London: Academic Press.

Deffenbacher, K.A., & Loftus, E.F. (1982). Do jurors share a common understanding of eyewitness behaviour? Law and Human Behaviour, 6, 15-30.

Dent, H.R. (1986). An experimental study of the effectiveness of different techniques of questioning mentally handicapped child witnesses. British Journal of Clinical Psychology, 25, 13-17.

Dhanens, T.P., & Lundy, R.M. (1975). Hypnotic and waking suggestions and recall. International Journal of Clinical and Experimental Hypnosis, 23, 68-79.

Diamond, B.L. (1988). Inherent problems in the use of pretrial hypnosis on a prospective witness. California Law Review, 68, 313-349.

Dietze, P.M., & Thomson, D.M. (1993). Mental Reinstatement of Context: A Technique for interviewing Child Witnesses. Applied Cognitive Psychology, 7, 97-108.

Diges, M (1988). Stereotypes and memory of real traffic accidents. In M.M. Gruneberg, P.E. Morris, R.N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 59-65 Chichester: John Wiley & Sons.

Dinges, D.F., Whitehouse, W.G., Orne, E.C., Powell, E.W., Orne, M.T., & Erdelyi, M.H. (1992). Evaluating Hypnotic Memory Enhancement (Hypermnesia and reminiscence) using Multitrial Forced Recall. Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 1139-1147.

Dorcus, R.M. (1960). Recall under hypnosis and amnesic events. International Journal of Clinical and Experimental Hypnosis, 8, 57-60.

Dunning, D., & Stern, L.B. (1992). Examining the Generality of Eyewitness Hypermnesia: A Close Look at Time Delay and Question Type. Applied Cognitive Psychology, 6, 643-657.

Dywan, J., & Bowers, K. (1983). The use of hypnosis to enhance recall. Science, 22, 184-185.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organisation of behaviour. Psychological Review, 66, 183-201.

Ebbinghaus, H. (1913). Memory. New York: Teacher's College Press.

Egeth, H.E. (1993). What Do We Not Know About Eyewitness Identification? American Psychologist, 48, 577-588.

Egeth, H.E., & McCloskey, M. (1984). The jury is still out: A reply to Deffenbacher. American Psychologist, 39, 1068-1069.

Elliott, R. (1993). Expert testimony about eyewitness identification Law and Human Behaviour, 17, 423-437.

Elliott, R., Farrington, B., & Manheimer, H. (1988). Eyewitnesses credible and discredible. Journal of Applied Social Psychology, 18, 1411-1422.

Erdelyi, M.E. (1994). Hypnotic hypermnesia: The empty set of hypermnesia. The International Journal of Experimental and Clinical Hypnosis, 42, 379-390.

Erdelyi, M.H. (1988). Hypermnesia: The effect of hypnosis, fantasy and concentration. In Pettinati H.M. (Eds), Hypnosis and Memory, pp. 64-94. New York: Guildford Press.

Erdelyi, M.H., Dinges, F.D., Orne, M.T., Whitehouse, W.G., & Orne, E.C., (1987). The stimulus and the test in hypnotic amnesia. Unpublished manuscript.

Erev, I., Wallsten, T.S., & Budescu, D.V. (1994). Simultaneous over- and underconfidence: The role of error in judgement processes. Psychological Review, 101, 519-527.

Evans, F.J. (1991). Hypnotisability: Individual differences in dissociation and the flexible control of psychological processes. In S.J. Lynn & J.W. Rhue (Eds). Theories of Hypnosis: Current Models and Perspectives. New York: Guildford.

Eysenck, H.J. (1941). An experimental study of the improvement of mental and physical functions in the hypnotic state. British Journal of Medical Psychology, 18, 304-316.

Faith, M., & Ray, W.J. (1994). Hypnotizability and dissociation in a college age population: Orthogonal individual differences.normal populations. Journal of Personality and Individual Differences, 17, 211-216.

Fischhoff, B, Slovic, S, & & Lichtenstein, S (1977). Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.

Fisher, R.P., & Chandler, C.C. (1991). Independence Between recalling Interevent Relations and Specific Events. Journal of Experimental Psychology: Learning, Memory and Cognition, 17, 722-733.

Fisher, R.P., & Geiselman, R.E. (1988). Enhancing eyewitness memory with the Cognitive Interview. In M.M. Gruneberg, P.E. Morris, R.N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 34-39. Chichester: John Wiley & Sons.

Fisher, R.P., & Geiselman, R.E. (1992). Memory enhancing techniques for investigative interviewing: The cognitive interview. Springfield: Charles C. Thomas.

Fisher, R.P., Geiselman, R.E., & Amador, M. (1989). Field Test of the Cognitive Interview: Enhancing the Recollection of Actual Victims and Witnesses of Crime. Journal of Applied Psychology, 74, 722-727.

Fisher, R.P., Geiselman, R.E., & Raymond, D.S. (1987). Critical Analysis of Police Interview Techniques. Journal of Police Science & Administration, 15, 177-185.

Fisher, R.P., Geiselman, R.E., Raymond, D.S., Jurkevich, L.M., & Warhaftig, M.L. (1987). Enhancing Enhanced Eyewitness Memory: Refining the Cognitive Interview. Journal of Police Science & Admintistration, 15, 291-297.

Fisher, R.P., & Price-Rouche, J. (1986). Question order and eyewitness memory. Unpublished Manuscript.

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The Effect of a five month delay on children's and adults' eyewitness memory. British Journal of Psychology, 83, 323-336.

Foster, R.A., Libkuman, T.M., & Schooler J.W. & Loftus, E.F. (1994). Consequentiality and eyewitness person identification. Applied Cognitive Psychology, 8, 107-121.

Fox, D.R. (1988). Psychological juriprudence and radical social change. American Psychologist, 48, 234-241.

Fox, S.G., & Walters, H.A. (1986). The impact of general versus specific expert testimony and eyewitness confidence upon mock juror judgement. Law and Human Behaviour, 10, 215-228.

Fruzzetti, A.E., Tolland, K., Teller, S.A., & Loftus, E.F. (1992). Memory and eyewitness testimony. In M.M. Gruneberg & P.E. Morris (Eds) Aspects of Memory: The Practical Aspects, pp. 18-50.London: Academic Press.

Garry, M., & Loftus, E.F. (1994). Pseudomemories without hypnosis. The International Journal Clinical and Experimental Hypnosis, 42, 363-378.

Geiselman, R.E. (1988). Improving eyewitness memory through mental reinstatement of context. In G.M. Davies & D.M. Thomson (Eds), Memory in Context: Context in Memory, pp. 245-266. Chichester: John Wiley & Sons Ltd.

Geiselman, R.E., Fisher, R.P., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness Response to Leading and Misleading Questions Unde the Cognitive Interview. Journal of Police Science & Administration, 14, 31-39.

Geiselman, R.E., Fisher, R.P., Firstenberg, I., Hutton, L.A., Sullivan, S.J., Avetissian, I.V., & Prosk, A.L. (1984). Enhancement of Eyewitness Memory: An Empirical Evaluation of the Cognitive Interview. Journal of Police Science and Administration, 12, 74-80.

Geiselman, R.E., Fisher, R.P., MacKinnon, D.P., & Holland, H.L. (1985). Eyewitness memory enhancement in the Police interview: Cognitive retrieval mnemonics versus hypnosis. Journal of Applied Psychology, 70, 401-402.

Geiselman, R.E., & Padilla, J. (1988). Interviewing child witnesses with the Cognitive Interview. Journal of Police Science and Administration, 16, 236-242.

George, R. (1990). A field and experimental evaluation of three methods of interviewing witnesses and victims of crime. Unpublished Masters Thesis, Polytechnic of East London, London, UK.

Gheorghi, V. (1972). Experimentalle unter suchugen zur hypnotischen hypermnesie. In D. Langen (Ed.), Hypnose und Psychosomatiche medizin, pp. 42-46. Stuttgart: Hippokrates Verlag.

Gibling, F., & Davies, G. (1988). Reinstatement of context following exposure to post-event information. British Journal of Psychology, 79, 129-141.

Glever, L. (1993). Why professionals lie. Organisational Behavior & Human, 55, 251-272.

Godden, D.R., & Baddeley, R. (1975). Context-dependent memory in two natural enviroments: on land and underwater. British Journal of Psychology, 66, 325-332.

Goodwin, D.W., Powell, B., Bremer, D., Hoine, H., & Stern, J. (1969). Alcohol and recall: State dependant effects in man. Science, 163, 1358.

Gorenstein, G.W., & Ellsworth, P.C. (1980). Effect of choosing an incorrect photograph on a later identification by an eyewitness. Journal of Applied Psychology, 65, 616-622.

Graffin, N.F., Ray, W., & Lundy, R. (1995). EEG concomitants of hypnosis and hypnotic susceptibility. Journal of Abnormal Psychology, 104, 123-131.

Green, D.M., & Swets, J.A. (1966). Signal Detection Theory and Psychophysics. New York: McGraw-Hill.

Gregg, V.H., & Mingay, D.J. (1987). Influence of hypnosis on riskiness and discriminability in recognition memory for faces. British Journal of Experimental and Clinical Hypnosis, 4, 65-75.

Gruneberg, M.M., & Sykes, R.N. (1993). The generalisability of confidence-accuracy studies in eyewitnessing Memory, 1, 185-189.

Gudjonsson, G. (1992). The Psychology of Interogations, Confessions and Testimony. Chichester: Wiley.

Gudjonsson, G. (1994). Confessions made to the expert witness: Some professional issues. Journal of Forensic Psychiatry, 5, 237-247.

Hall, D.F., Loftus, E.F., & Tousignant, J.P. (1984). Postevent information and changes in recollection for a natural event. In Wells G.L. & Loftus E.F.

(Eds), Eyewitness Testimony: Psychological Perspectives, pp. 124-141. New York: Cambridge University Press.

Hart, J.T. (1967). Memory and the memory-monitoring process. Journal of Verbal Learning and Verbal Behaviour, 6, 685-691.

Hart, J.T. (1966). Methodological note on feeling of knowing experiments. Journal of Educational Psychology, 57, 347-349.

Hart, J.T. (1965). Memory and the feeling of knowing experience. Journal of Educational Psychology, 56, 208-216.

Hastie, R., Landsman, R., & Loftus, E.F. (1978). Eyewitness testimony: The dangers of guessing Jurimetrics Journal, 19, 1-8.

Hastorf, A.H., & Cantril, H. (1954). They saw the game: A case study. Journal of Abnormal and Social Psychology, 49, 129-134.

Haward, L., & Ashworth, A. (1980). Some problems of evidence obtained by hypnosis. Criminal Law Review, 469-485.

Herman, D., & Gruneberg, M. (1993). The need to expand the horizons of the practical aspects of memory movement. Applied Cognitive Psychology, 7, 553-565.

Herren, R. (1976). Das Vernehmungsprotokoll. Kriminalistik, 7, 313-317.

Hibbard, W.S., & Worring, R.W. (1981). Forensic hypnosis: The practical application of hypnosis in criminal investigation. Springfield, Illonois, USA: C.C. Thomas.

Hilgard, E.R. (1991). A neodissociation interpretation of hypnosis. In S.J. Lynn & J.W. Rhue (Eds) Theories of Hypnosis: Current Models and Perspectives. New York: Guildford.

Hilgard, E.R. (1986). Divided Consciousness: Multiple Controls in Human Thought and Action. New York: Wiley.

Hintzman, D.L. (1976). Repetition and memory. In G.L. Bower (Ed.) The Psychology of Learning and Motivation. New York: Academic Press.

Hollin, C.R. (1989). Criminological and legal psychology. Psychology Survey, 7, 146-175.

Hollin, C.R., & Clifford, B.R. (1983). Eyewitness testimony: The effects of discussion on recall accuracy and agreement. Journal of Applied Social Psychology, 13, 234-244.

Holst, V.F., & Pezdek, K. (1992). Scripts for Typical Crimes and their Effects on Memory for Eyewitness Testimony. Applied Cognitive Psychology, 6, 573-587.

Hosch, H.M., & Cooper, D.S. (1982). Victimization as a determinant of eyewitness accuracy. Journal of Applied Psychology, 67, 649-652.

Howell, D.C. (1992). Statistical Methods for Psychology, 3rd Ed. Belmon, California: Duxby Press.

Huff, C.R., & Rattner, A. (1988). Convicted but innocent: False positives and the criminal justice process. In E. Scott & T. Hirschi (Eds) Controversial issues in crime and justice. Beverley Hills, California: Sage.

Huse, B. (1930). Does the hypnotic trance favor the recall of faint memories? Journal of Experimental Psychology, 13, 519-529.

Hyman, I.E. (1994). Conversational remembering: Story recall with a peer versus for an experimenter. Applied Cognitive Psychology, 8, 49-66.

Isen, A.M., Shalker, T.E., Clark, M., & Karp, L. (1978). Affect, Accessibility of Material in Memory, and Behaviour: A Cognitive Loop? Journal of Personality and Social Psychology, 36, 1-12.

Jobe, J.B., Tourangeau, R., & Smith, A.F. (1993). Contributions of survey research to the understanding of memory. Applied Cognitive Psychology, 7, 567-584.

Johnson, M.T. (1993). Memory phenomena in the law. Applied Cognitive Psychology, 7, 603-618.

Joordens, S., & Merikle, P.M. (1992). False recognition and perception without awareness. Memory and Cognition, 20, 151-159.

Kassin, S.M., & Ellsworth P.C. & Smith, V.L. (1994). Deja vu all over again: Elliot's critique of eyewitness experts. Law and Human Behaviour, 18, 203-210.

Kassin, S.M., Ellsworth, P.C., & Smith, V.L. (1989). The 'general acceptance' of pychological research on eyewitness testimony. American Psychologist, 44, 1089-1098.

Kelman, H.C. (1967). Human Use of Subjects: The problem of deception in social psychological experiments. Psychological Bulletin, 67, 1-11.

Kihlstrom, J.F. (1985). Hypnosis. Annual Review of Psychology, 36, 385-418.

Kihlstrom, J.F. (1994). Hypnosis, delayed recall, and the principles of memory. The International Journal of Clinical and Experimental Hypnosis, 42, 337-345.

Kinnunen, T., Zamansky, H.S., & Block, B.L. (1994). Is the hypnotized subject lying? Journal of Abnormal Psychology, 103, 184-191.

Kirsch, I. (1992). The state of the altered state debate. Contemporary Hypnosis, 9, 1-6.

Klatzky, R.L., & Erdelyi, M.H. (1985). The response criterion problem in tests of hypnosis and memory. Journal of Clinical and Experimental Hypnosis, 33, 246-257.

Kleinhauz, M., Horrowitz, I., & Tobin, T. (1977). The use of hypnosis in police investigation: A preliminary communication. Journal of the Forensic Science Society, 17, 77-80.

Kohnken, G., & Brockman, C. (1987). Unspecific postevent information, attribution of resposibility, and eyewitness performance. Applied Cognitive Psychology, 1, 197-207.

Kohnken, G., Finger, M., Nitschke, N., Hofer, E., & Ascherman, E. (1992). Paper presented at the Conference of the American Psychology-Law Society, San Diego, USA.

Kohnken, G., & Maas, A. (1988). Eyewitness Testimony: False Alarms on Biased instructions? Journal of Applied Psychology, 73, 363-370.

Kohnken, G., Thurer, C., & Zoberier, D. (1994). The cognitive interview: Are the interviewer's memories enhanced, too? Applied Cognitive Psychology, 8, 13-24.

Konecni, V.J., & Ebbeson, E.B. (1986). Courtroom testimony by psychologists on eyewitness identification issues: Critical notes and reflections. Law and Human Behaviour, 10, 117-126.

Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. Psychological Review, 100, 609-639.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy orientated and quantity-orientated approaches to memory assessment. Journal of Experimental Psychology: General, 123, 297-315.

Koslowsky, M. (1993). The Use of Power Tactics to Gain Compliance. Social Behaviour & Personality, 21, 135-143.

Kossowska, M., & Necka, E. (1994). Do it your own way: Cognitive stratergies, intelligence, an personality. Personality and Individual Differences, 16, 33-46.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. Journal of Personality and Social Psychology, 49, 58-69.

Kramer, T.H., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal and eyewitness memory: Attention must be paid. Law and Human Behaviour, 10, 117-126.

Kroger, W.S. (1979). Hypnosis in clinical investigation The International Journal of Clinical and Experimental Hypnosis, 27, 258-274.

Kuehn, L.L. (1974). Looking down a gun barrel: Person perception and violent crime. Perceptual and Motor Skills, 39, 1159-1164.

Kuzins, R. (1989). False accusations. L.A. Lawyer, Sept, 58-59.

Laughery, K.R., Alexander, J.F., & Lane, A.R. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position and type of photograph. Journal of Applied Psychology, 55, 477-483.

Leippe, M.R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. Law and Human Behaviour, 4, 261-274.

Leippe, M.R., Maniou, A.P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? Journal of Personality and Social Psychology, 63, 181-197.

Leippe, M., Wells, G.L., & Ostrom, T. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. Journal of Applied Psychology, 3, 345-351.

Lepore, S.J., & Sesco, B. (1994). Distorting children's reports and interpretations of events through suggestion. Journal of Applied Psychology, 79, 108-120.

Lindsay, R.C.L. (1994). Expectations of eyewitness performance: Jurors verdicts do not follow from their beliefs. In D.F. Ross, J.D. Read & M.P. Toglia (Eds). Adult Eyewitness Testimony: Current trends and developments, pp. 362-382 New York: Cambridge University Press.

Lindsay, R.C.L., & Harvis, V.L. (1988). Hits, false alarms, correct and mistaken identifications: The effects of method of data collection on facial memory. In M.M. Gruneberg, P.E. Morris, R.N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 47-52 Chichester: John Wiley & Sons.

Lindsay, R.C.L., Nosworthy, G.J., Martin, R., & Martynuck, C. (1994). Using mug shots to find suspects. Journal of Applied Psychology, 79, 121-130.

Linday, R.C.L., & Wallbridge, H. (1983). Do the clothes make the man? An exploration of the effect of line-up attire on eyewitness identification. Unpublished manuscript.

Lindsay, R.C.L., Wells, G.L., & O'Conner, F.J. (1989). Mock-juror belief or accurate and inaccurate eyewitnesses: A replication and extension. Law and Human Behaviour, 13, 333-339.

Lindsay, R.C.L., Wells, G.L., & Rumpel, C.M. (1981). Can people detect eyewitness-identification accuracy within and accross situations? Journal of Applied Psychology, 66, 79-89.

Loftus, E.F. (1972). Eye fixations and recognition memory. Cognitive Psychology, 3, 525-557.

Loftus, E.F. (1975). Leading questions and eyewitness report Cognitive Psychology, 7, 560-572.

Loftus, E.F. (1979). Reactions to blatantly contradictory information. Memory and Cognition, 7, 368-374.

Loftus, E.F. (1981). Eyewitness testimony: Psychological research and legal thought. In M. Tonry & N. Morris (Eds) Crime and Justice: An Annual Review of Research. Chicargo: University of Chicargo Press.

Loftus, E.F. (1992). When a lie becomes memory's truth: Memory distortion after exposure to misinformation. Current Directions in Psychological Science, 1, 121-123.

Loftus, E.F. (1993a). The reality of repressed memories. American Psychologist, 48, 518-537.

Loftus, E.F. (1993b). Psychologists in the Eyewitness World. American Psychologist, 48, 550-552.

Loftus, E.F., Altman, D., & Geballe, R. (1975). Effects of questioning of a witness's later recollections. Journal of Police Science and Administration, 3, 162-165.

Loftus, E.F., & Burns, T.E. (1982). Mental shock can produce retrograde amnesia. Memory and Cognition, 10, 318-323.

Loftus, E.F., & Greene, E. (1980). Warning: Even memory for faces can be contageous. Law and Human Behaviour, 4, 323-334.

Loftus, E.F., Loftus, G.R., & Messo, J. (1987). Some facts about weapon focus. Law and Human Behaviour, 11, 55-62.

Loftus, E.F., & Palmer, J.C. (1974). Reconstruction of automobile distruction: An example of interaction between language and memory. Journal of Verbal Learning and Verbal Behaviour, 13, 585-589.

Loftus, E.F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. Bulletin of the Psychonomic Society, 5, 86-88.

Lonigan, C.J., Shannon, M.P., Taylor, C.M., Finch, A.J., & Sallee, F.R. (1994). Children exposed to disaster: II. Risk factors for the development of post-traumatic syptomatology. Journal of the American Academy of Child and Adolescent Psychiatry, 33, 94-105.

Lore, R.K., & Schultz, L.A. (1993). Control of human agression: A comparative perspective. American Psychologist, 48, 16-25.

Lundeberg, M.A., Fox, P.W., & Puncochar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. Journal of Educational Psychology, 86, 114-121.

Luus, C.A., & Wells, G.L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. Journal of Applied Psychology, 79, 714-723.

Lynn, S.J., & Rhue, J.W. (1991). In (Eds), Theories of Hypnosis: Current Models and Perspectives. New York: Guilford.

Lynn, S.J. (1992). A non-state view of hypnotic involutariness. Contempory Hypnosis, 9, 21-27.

Maass, A., & Kohnken, G. (1989). Eyewitness identification: Simulating the 'witness effect'. Law and Human Behaviour, 11, 397-408.

Malpass, R.S., & Devine, P.G. (1981). Eyewitness Identification: Lineup Instructions and the Absence o the Offender. Journal of Applied Psychology, 66, 482-489.

Mantwill, M., Kohnken, G. (In press). Effects of the cognitive interview on the recall for familiar and unfamiliar events. Journal of Applied Psychology

Marquis, K.H., Marshall, J., & Oskamp, S. (1972). Testimony Validity as a Function of Question Form, Atmosphere, and Item Difficulty. Journal of Applied Social Psychology, 2, 167-186.

McCarthy, K. (1993). Kids' eyewitness recall is focus for conference. The American Psychological Association, 24(7), 1 & 28.

McCloskey, M., Egeth, H.E., & McKenna, J. (1986). The experimental psychologist in court: The ethics of expert testimony. Law and Human Behaviour, 10, 1-13.

McConkey, K.M., & Nogrady, H. (1984). Hypnosis, hypnotizability, and story recall. Australian Journal of Clincal and Experimental Hypnosis, 12, 93-98.

McConkey, K.M., & Roche, S.M. (1989). Knowledge of eyewitness memory. Australian Psychologis, 24, 377-384.

McEwan, N.H., & Yuille, J.C. (1981). The effects of training and experience on eyewitness memory. Paper presented to the Canadian Psychological Association, Canada.

McSpadden, M.D., Schooler, J.W., & Loftus, E.F. (1988). Here today, gone tomorrow: The appearance and disappearance of context effects. In G.M. Davies & D.M. Thomson (Eds), Memory in Context: Context in Memory, pp. 215-229. Chichester: John Wiley & Sons Ltd.

Memon, A. (1991). Introducing the cognitive interview as a procedure for interviewing witnesses. Paper presented to the Annual conference of the British Psychological Society, Bournmouth, UK.

Memon, A., & Bull, R. (1991). The Cognitive Interview: its Origins, Empirical Support, Evaluation and Practical Implications. Journal of Community & Applied Social Psychology, 1, 291-307.

Memon, A., Bull, R., & Smith, M. (1995). Improving the quality of the police interview: Can training in the use of the cognitive techniques help? Policing and Society, 5, 53-68.

Memon, A., Cronin, O., Eaves, R., and Bull, R. (1993). The cognitive interview and child witnesses. Paper presented the the Division of Criminological Psychology, Harrogate, UK.

Memon, A., Holley, A., Milne, R, Kohnken, G., & Bull, R. (1994). Towards understanding the effects of interviewer training in evaluating the cognitive interview. Applied Cognitive Psychology, 8, 641-659.

Memon, A., & Kohnken, G. (1992). Helping witnesses to remember more: the cognitive interview. Expert Evidence,2, 39-48.

Memon, A., Wark, L., Bull, R., & Kohnken, G. (in press). Children's event memory: the effects of interview technique and repeated testing. Journal of Experimental Psychology

Metcalfe, J. (1990). Composite Holographic Associative Recall Model (CHARM) and blended Memories in Eyewitness Testimony. Journal of Experimental Psychology: General, 119, 145-160.

Meudel, P.R., Hitch, G.J., & Kirby, P. (1992). Are Two Heads Better than One? Experimental Investigations o the Social Facilitation of Memory Applied Cognitive Psychology, 6, 525-543.

Mitchell, M.B. (1932). Retroactive inhibition and hypnosis. Journal of General Psychology, 7, 343-359.

Miller, D.G., & Loftus, E.F. (1976). Influencing memory for people and their actions. Bulletin of the Psychonomic Society, 7, 9-11.

Milne, R., & Bull, R. (1994). Improving witness recall: The Cognitive Interview and the Legal profession. Journal of Child Law, 6, 82-84.

Milne, R., Bull, R., Koehnken, G., & Memon, A. (1994). The Cognitive Interview and child witness suggestibility. Poster presentation to the British Pschological Society Annual Confernce, Brighton, UK.

Mingay, D. (1987). The effect of hypnosis on eyewitness memory: Reconcilling forensic claims and research findings. Applied Psychology: An International Review, 36, 163-183.

Mingay, D.J. (1986). Hypnosis and memory for incidentally learned scenes. British Journal of Experimental and Clinical Hypnosis, 3, 173-183.

Murrey, G.J., Cross, H.J., & Whipple, J. (1992). Hypnotically created pseudomemories: Further investigation into the "memory distortion or response bias" question. Journal of Abnormal Psychology, 101, 75-77.

Nash, M.R. (1991). Hypnosis as a special case of psychological regression. In S.J. Lynn & J.W. Rhue (Eds), Theories of Hypnosis: Current Models and Perspectives. New York: Guildford.

Nash, M.R., Drake, S.D., Wiley, S., Khalsa, S., & Lynn, S.J. (1986). Accuracy of Recall by Hypnotically Age-Regressed Subjects. Journal of Abnormal Psychology, 95, 298-300.

Nasj, M.R. (1994). Memory distortion and sexual trauma: The problem of false negatives and positives. The International Journal of Clinical and Experimental Hypnosis, 42, 346-362.

Nelson, T.O. (1988). A comparison of current measures of the accuracy of feeling of knowing predictions. Psychological Bulletin, 95, 109-133.

Nelson, T.O., Gerler, D., Narens, L. (1984). Accuracy of feeling of knowing judgements for predicting perceptual identification and relearning. Journal of Experimental Psychology- General, 113, 282-300.

Newlands, P., & George, R. (1994). Are police officers consistent in their judgements of what is a 'good' witness account? Paper presented to the Annual Conference of the Psychology Postgraduate Affairs Group, Sheffield, UK.

Nicholas, L.J., Durrheim, K., & Tredoux, C.G. (1994). Lying as a factor in research on sexuality. Psychological Reports, 75, 839-842.

Nightingale, G., Kebbell, M.R., Thomas, S., & Wagstaff, G.F. (1995). The role of cognitive interview and fear on the recall of first time parachutists. Poster presentation to the Annual Conference of the British Psychological Society, Warwick, UK..

Nogrady, H., McConkey, & Perry, C. (1985). Enhancing Visual Memory: Trying Hypnosis, Trying Imagination and Trying Again. Journal of Abnormal Psychology, 94, 195-204.

Noon, E., & Hollin, C.R. (1987). Lay knowledge of eyewitness behaviour: A British survey. Applied Cognitive Psychology, 1, 143-153.

O'Rourke, T.E., Penrod, S.D., Cutler, B.L., & Stuve, T.E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. Law and Human Behaviour, 13, 385-395.

Ofshe, R.J., & Singer, M.T. (1994). Recovered-memory therapy and robust repression: Influence and pseudomemories. The International Journal of Clinical and Experimental Hypnosis, 4, 391-410.

Orne, M.T. (1962). On the Social Psychology of the Psychological Experiment: With particular reference to demand charecteristics and their implications. American Psychologist, 17, 776-783.

Orne, M.T., Dinges, D.F., & Orne, E.C. (1986). Hypnotic experience: A cognitive social psychological reality. Behavioural and Brain Sciences, 9, 477-478.

Orne, M.T., & Evans, F.J. (1965). Social Control in the Psychological Experiment Journal of Personality and Social Research, 1, 189-200.

Orne, M.T., Sokis, D.A., Dinges, D.F., & Orne, E.C. (1984). Hypnotically induced testimony. In G.L. Wells & E.F. Loftus (Eds), Eyewitness Testimony: Psychological Perspectives, pp. 171-215 Cambridge, England: Cambridge University Press.

Orne, M.T. (1979). The use and misuse of hypnosis in court. The International Journal of Clinical and Experimental Hypnosis, 27, 311-341.

Orne, M.T., Whitehouse, W.G., Dinges, D.F., & Orne, E.C. (1988). Reconstructing memory through hypnosis: Forensic and clinical implications. In Pettinati H.M. (Eds), Hypnosis and Memory, pp. 21-63 New York: Guildford Press.

Perfect, T.J., Watson, E.L., & Wagstaff, G.F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. Journal of Applied Psychology, 78, 144-147.

Perlman, N.B., Ericson, K.I., & Esses V.M. & Isaacs, B.J. (1994). The developmentally handicapped witness. Law and Human Behaviour, 18, 171-187.

Perry, C. (1979). Hypnotic coercion and compliance to it: A review of evidence presented in a legal case. The International Journal of Clinical & Experimental Hypnosis, 27, 187-218.

Perry, C.W., Laurence, J.R., D'Eon, J., & Tallant, B. (1988). Hypnotic age regression techniques in the elicitation of memories: Applied uses and abuses. In Pettinati H.M. (Eds), Hypnosis and Memory, pp. 128-154 New York: Guildford Press.

Perry, C.W., & Nogrady, H. (1985). Use of hypnosis in the investigation of crime: is guided memory a safe substitute? British Journal of Experimental and Clinical Hypnosis, 3, 25-31.

Peters, D.P. (1988). Eyewitness memory and arousal in a natural setting. In M. M. Gruneberg, P. E. Morris, R. N. Sykes (Eds), Practical Aspects of Memory: Current Research and Issues, pp. 89-94 Chichester: John Wiley & Sons.

Peterson, M.A. (1976). Witness: Memory of social events. Unpublished PhD thesis, University of California, Los Angeles, USA.

Putman, W.H. (1979). Hypnosis and distortions in eyewitness memory. The International Journal of Clinical & Experimental Hypnosis, 27, 437-448.

Radke, H.L., & Spanos, N.P. (1981). Was I hypnotised?: A social psychological analysis of hypnotic depth reports. Psychiatry, 44, 359-376.

Rainer, D. (1983). Eyewitness testimony: Does hypnosis enhance accuracy, distortion and confidence? Unpublished PhD thesis, University of Wyoming, USA.

Rand, Corporation (1975). The Criminal Investigation Process Rand Corporation Technical Report R-1777, 1-3.

Read, J.D., Tollestrup, P., Hammersly, R., McFadzen, E., & Christensen, A. (1990). The unconcious transference effect: Are innocent bystanders ever misidentified? Applied Cognitive Psychology, 4, 3-31.

Redston, M.T., & Knox, J. (1983). Is the recognition of faces enhanced by hypnosis? Paper presented to the Annual Meeting of the Society of Clinical and Experimental Hypnosis, Boston, USA.

Register, P.A., & Kihlstrom, J.F. (1987). Hypnotic effects on hypermnesia. Internation Journal of Clinical and Experimental Hypnosis, 35, 155-170.

Reiser, M. (1980). Handbook of Investigative Hypnosis. Los Angeles: Lehi..

Reiser, M. (1990) cited In Gudjonsson, G. (1992). The Psychology of Interogations, Confessions and Testimony. Chichester: Wiley.

Rosenthal, B.G. (1944). Hypnotic recall of material learned under anxiety and non-anxiety producing conditions. Journal of Experimental Psychology, 34, 369-389.

Ross, L., Lepper, M.R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 32, 880-892.

Salzberg, H.C., & DePiano, F.A. (1980). Hypnotizability and task motivating instructions: A further look at how they effect performance. International Journal of Experimental and Clinical Hypnosis, 28, 261-271.

Sanders, G.S. (1984). Effects of Context Cues on Eyewitness Identification Responses. Journal of Applied Social Psychology, 14, 386-397.

Sanders, G.S. (1986). The Usefulness of Eyewitnesses from the Perspective of Police Investigators. Unpublished Manuscript, State University of New York at Albany, USA.

Sanders, G.S., & Simmons, W.L. (1983). Use of hypnosis to enhance eyewitness accuracy: Does it work? Journal of Applied Psychology, 68, 70-77.

Sarbin, T.R., & Coe, W.C. (1972). Hypnosis: A Social Psycholocal Analysis of Influence Communication. New York: Holt, Rinehart and Winston.

Saywitz, K.J., Geiselman, R.E., & Bornstein, G.K. (1992). Effects of cognitive interviewing and practice on children's recall performance Journal of Applied Psychology.

Schacter, D.L. (1983). Feeling of knowing in episodic memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 39-54.

Schmidt, M.I., Duncan, B.B., Tavares, M., Polanczyk, Pellanda, L., & Zimmer, P.M. (1993). Validity of self-reported weight. A study of urban Brazilian adults. Revista de Saude Publica, 27, 271-276.

Scogin, F., Calhoon, S.K., & D'Errico, M. (1994). Eyewitness confidence and accuracy among three age cohorts. Journal of Applied Gerontology, 13, 172-184.

Sears, A.B. (1954). A comparison of hypnotic and waking recall. Journal of Clinical and Experimental Hypnosis, 2, 296-304.

Shames, M.L. (1981). Hypnotic susceptibility and conformity: On the mediational mechanism of suggestibility. Psychological Reports, 49, 563-566.

Shannon, M.P., Lonigan, C.J., Finch, A.J., & Taylor, C.M. (1994). Children exposed to disaster: I. Epidemiology of post-traumatic symptoms and symptom profiles. Journal of the American Acadamy of Child and Adolescent Psychiatry, 33, 80-93.

Shaw, J.I., & Skolnick, P. (1994). Sex differences, weapon-focus, and eyewitness reliability. The Journal of Social Psychology, 134, 413-420.

Sheehan, P.W. (1988). Confidence, memory and hypnosis. In H.M. Pettinati (Ed), Hypnosis and Memory, pp. 95-127 New York: Guildford Press.

Sheehan, P.W., Grigg, L., & McCann, T. (1984). Memory distortion following exposure to false information in hypnosis. Journal of Abnormal Psychology, 93, 259-296.
Sheenhan, P.W., Green, V., & Truesdale, P. (1992). Influence of rapport on hypnotically induced pseudomemory Journal of Abnormal Psychology, 101, 690-700.

Sheehan, P.W., & McConkey, K.M. (1982). Hypnosis and Experience: The Exploration of Phenomena and Process Hillsdale: Erlbaum.

Sheehan, P.W., & Tilden, J. (1983). Effects of suggestibility and hypnosis on accurate and distorted retrieval from memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 283-293.

Sheehan, P.W., & Tilden, J. (1984). Real and simulated occurrences of memory distortion in hypnosis. Journal of Abnormal Psychology, 93, 47-55.

Sheehan, P.W., & Tilden, J. (1986). The consistency of occurrences of memory distortion following hypnotic induction. International Journal of Clinical and Experimental Hypnosis, 2, 122-137.

Smith, J.M., Bell, P.A., & Fusco, M.E. (1986). Influence of Colour and Demand Charecteristics on Muscle Strengt and Affective Ratings of the Enviroment The Journal of General Psychology, 113, 289-297.

Smith, M.C. (1983). Hypnotic memory enhancement of witnesses: Does it work? Psychological Bulletin, 94, 387-407.

Smith, S.M. (1979). Remembering in and out of context. Journal of Experimental Psychology: Human Learning & Memory, 5, 460-471.

Smith, S.M., Glenberg, A., & Bjork, R.A. (1978). Enviromental Context & Human Memory. Memory & Cognition, 6, 342-253.

Smith, V.L., Kasin, S.M., & Ellsworth, P.C. (1989). Eyewitness accuracy and confidence: Within versus between-subjects correlations Journal of Applied Psychology, 74, 356-359.

Spanos, N.P. (1982). A social psychological approach to hypnotic behaviour Integration of Clinical and Social Psychology, , 231-271.

Spanos, N.P. (1986). Hypnotic behaviour: A social psychological interpretation of amnesia, analgesia and 'trance logic'. The Behavioural and Brain Sciences, 9, 449-502.

Spanos, N.P. (1991). A sociocognitive approach to hypnosis. In S.J. Lynn & J.W. Rhue (Eds), Theories of Hypnosis: Current Models and Perspectives. New York: Guilford.

Spanos, N.P. (1992). Compliance and reinterpretation in hypnotic responding Contempory Hypnosis, 9, 7-15.

Spanos, N.P., & Chaves, J.F. (1989). Hypnosis: The Cognitive Behavioural Perspective. New York: Guilford.

Spanos N.P., Gwynn M.I., Comer S.L. & Baltruweit W.J. & de Groh, M. (1989). Are hypnotically induced pseudomemories resistant to cross-examination? Law & Human Behaviour, 13, 271-289.

Spanos, N.P., Gwynn, M.I., & Terrade, K. (1989). Effects on mock jurors of experts favourable and unfavourable toward hypnotically elicited eyewitness testimony. Journal of Applied Psychology, 74, 922-926.

Spanos N.P., Quigley C.A., Gwynn M.I., Glatt R.L., & Perlini A.H. (1991). Hypnotic interrogation, pretrial preparation, and witness testimony during direct and cross-examination. Law and Human Behaviour, 25, 639-653.

Sporer, S.L. (1983). Allgemeinwissen zur psychologie der zeugenaussage.In H.J. Kerner, H. Kury & K. Sessar (Eds), Deutche Forschungen zur Kriminalitaetsentstehung und Kriminalitaetskontrolle, pp. 1191-1234. Koeln: Heymanns.

Sporer, S.L. (1992). Post-dicting eyewitness accuracy: Confidence, descision times and person descriptions of choosers and non-choosers. European Journal of Social Psychology, 22, 157-180.

Sporer, S.L. (1993). Eyewitness identification accuracy: Confidence, decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78, 22-33.

Sporer, S.L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D.F. Ross, J.D. Read & M.P. Toglia (Eds), Adult eyewitness testimony: Current trends and developments, pp. 300-327. New York: Cambridge University Press.

Sporer, S.L., Penrod, S., Read, D., & Cutler, B. (in press). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. Psychological Bulletin.

Stager, G.L., & Lundy, R.M. (1985). Hypnosis and the learning and recall of visually presented material. International Journal of Clinical and Experimental Hypnosis, 33, 27-39.

Stalnaker, J.M., & Riddle, E.E. (1932). The effect of hypnosis on long delayed recall. Journal of General Psychology, 6, 429-440.

Stam, H.J., & Spanos, N.P. (1980). Experimental design, expectancy effects, and hypnotic analgesia. Journal of Abnormal Psychology, 89, 751-762.

Steblay, N.M., & Bothwell, R.K. (1994). Evidence for hypnotically refreshed testimony. Law & Human Behaviour, 18, 635-651.

Stephenson, G.M. (1984). Accuracy and confidence in testimony: A critical review of some fresh evidence. In In D.J. Muller, D.E. Blackman, & A.J. Chapman (Eds), Psychology and Law: Topics from an international conference, pp. 229-248 Chichester: Wiley.

Stephenson, G.M., Brandstatter, H., & Wagner, W. (1983). An experimental study of social performance and delay on the testimonial validity of story recall. European Journal of Social Psychology, 13, 175-1191.

Stephenson, G.M., Clark, N.K., & Wade, G.S. (1986). An experimental study of collaborative and individual recall of a simulated police interrogation. Journal of Personality and Social Psychology, 50, 1113-1122.

Tart, C.T. (1970). Self report scales of hypnotic depth. International Journal of Clinical and Experimental Hypnosis, 18, 105-125.

Taylor, L., Zuckerman, B., Harik, V., & McAlister Groves, B. (1994). Witnessing violence by young children and their mothers. Developmental and Behavioural Pediatrics, 15, 120-123.

Teasdale, J.D., & Fogrty, S.J. (1979). Differential effects of induced mood on retrieval of pleasent and unpleasant events from episodic memory. British Journal of Clinical Psychology, 22, 248-257.

Timm, H.W. (1981). The effect of forensic hypnosis techniques of eyewitness recall and recognition. Journal of Police Science and Administration, 9, 188-194.

Tippett, V. (1993). An investigation into the effects of hypnosis on memory. Unpublished manuscript, University of Liverpool, Liverpool, UK.

Tooley, V., Brigham, J.C., Maass, A., & Bothwell, R.K. (1987). Facial recognition: Weapon effect and attentional focus. Journal of Applied Social Psychology, 17, 845-859.

United Kingdom Police Central Planning and Training Unit (1992). Investigative Interviewing: A guide to interviewing. Unpublished manuscript.

Wagstaff, G.F. (1981a). Hypnosis, Compliance and Belief Brighton: Harvester.

Wagstaff, G.F. (1981b). The use of hypnosis in police investigation Journal of the Forensic Science Society, 21, 3-7.

Wagstaff, G.F. (1982a). Hypnosis and recognition of a face. Perceptual and Motor Skills, 55, 816-818.

Wagstaff, G.F. (1982b). Hypnosis and witness recall: A discussion paper. Journal of the Royal Society of Medicine, 75, 793-797.

Wagstaff, G.F. (1982c). Recall of witnesses under hypnosis. Journal of the Forensic Science Society, 22, 33-39.

Wagstaff, G.F. (1984). The enhancement of witness memory by hypnosis: A review and methodological critique of the experimental literiture. British Journal of Experimental and Clinical Hypnosis, 2, 3-12.

Wagstaff, G.F. (1985). The use of hypnosis by the police in the investigation of crime: Is guided imagery a safe substitute? British Journal of Experimental and Clinical Hypnosis, 3, 39-42.

Wagstaff, G.F. (1986). State versus non-state paradigms of hypnosis: a real or false dichotomy? Behavioural and Brain Sciences, 9, 486-487.

Wagstaff, G.F. (1989). Forensic Aspects of Hypnosis. In N.P. Spanos & J.F. Chaves (Eds) Hypnosis the cognitive behavioural perspective Buffalo, New York: Prometheus.

Wagstaff, G.F. (1991). Compliance, belief and semantics in hypnosis: A non-state socio-cognitive perspective. In S.J. Lynn & J.W. Rhue (Eds), Theories of Hypnosis: Current Models and Perspectives. pp. New York: Guilford.

Wagstaff, G.F. (1993). What expert witnesses can tell courts about hypnosis: A review of the associarion between hypnosis and the law Expert Evidence, 2, 60-70.

Wagstaff, G.F. (1995). Hipnosis y psicologia forense. In A. Capafons, S. Amigo, & I. Kirsch (Eds), Cognitive Behavioural Hypnosis and Self-Regulation Therapy. Valen, Spain: Promolibro.

Wagstaff, GF, & Mercer, K (1993). Does hypnosis facilitate memory for deep processed stimuli? Contempory Hypnosis, 10, 59-66.

Wagstaff, G.F., & Ovenen, M. (1979). Hypnotic time distortion and free recall learning- an attempted replication. Psychological Research, 40, 291-298.

Wagstaff, G.F., & Royce, C. (1994). Hypnosis and the treatment of nail biting: A preliminary trial. Contemporary Hypnosis, 11, 9-13.

Wagstaff, G.F., & Sykes, C.T. (1983). Hypnosis and the recall of emotionally toned material. IRCS Medical Science, 12, 137-138.

Wagstaff, G.F., Traverse, J., & Milner, S. (1982). Hypnosis and eyewitness memory: two experimental analogues. IRCS Medical Science, 12, 137-138.

Wagstaff, G.F., Vella, M., & Perfect, T.J. (1992). The effect of hypnotically elicited testimony on juror's judgements of guilt and innocence. Journal of Social Psychology, 31, 69-77.

Warner, K.E. (1979). The use of hypnosis in the defence of criminal cases. The International Journal of Clinical & Experimental Hypnosis, 27, 417-436.

Warnick, D.H., & Sander, G.S. (1980). The effects of group discussion on eyewitness accuracy. Journal of Applied Social Psychology, 10, 249-259.

Weber, S.J, & Cook, T.D. (1972). Subject effects in laboratory research: An examination of subject roles, demand charecteristics, and valid inference. Psychological Bulletin, 77, 273-295.

Wells, G.L. (1993). What Do We Know About Eyewitness Identification? American Psychologist, 48, 553-571.

Wells, G.L., & Elizabeth, C.A. (1990). The diagnosticity of a lineup should not be confused with the diagnostic-value of non-lineup evidence. Journal of Applied Psychology, 75, 511-516.

Wells, G.L., & Murray, M. (1984). Eyewitness confidence. In Wells G.L., Loftus E.F. (Eds), Eyewitness Testimony: Psychological Perspectives, pp. 155-170 New York: Cambridge University Press.

Wells, G.L. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy. Journal of Applied Psychology, 70, 619-626.

Wells, G.L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. Journal of Personality and Social Psychology, 36, 1546-1557.

Wells, G.L., Ferguson, T.J., & Lindsay, R.C.L. (1981). The tractability of eyewitness confidence and its implications for triers of fact. Journal of Applied Psychology, 66, 688-696.

Wells, G.L., & Leippe, M.R. (1981). How do triers of fact infer the accuracy of eyewitness identifications? Using memory for peripheral detail can be misleading. Journal of Applied Psychology, 66, 682-687.

Wells, G.L., Lindsay, R.C.L., & Ferguson, T.J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. Journal of Applied Psychology, 64, 440-448.

Wells, G.L., & Turtle, J.W. (1987). Eyewitness testimony research: Current knowledge and emergent contorversies. Canadian Journal of Behavioural Sciences, 19, 363-388.

Whipple, G.M. (1909). The observer as reporter: A survey of the 'psychology of testimony'. Psychological Bulletin, 6, 153-170.

White, R., Fox, G.R., & Harris, W.W. (1940). Hypnotic hypermnesia: Enhanced memory accessibility or report bias? Journal of Abnormal and Social Psychology, 35, 88-103.

Whitehouse, W.G., Dinges, D.F., Orne, E.C., & Orne, M. (1988). Hypnotic hypermnesia: Enhanced memory accessibility or report bias? Journal of Abnormal Psychology, 97, 289-295.

Whittlesea, W.A. (1987). Preservation of specific experiences in the representation of general knwledge. Journal of Experimental Psychology: Learning, Memory & Cognition, 13, 3-17.

Worthington, T.S. (1979). The use in court of hypnotically enhanced testimony. The International Journal of Clinical & Experimental Hypnosis, 27, 402-416.

Yarmey, A.D. (1986). Verbal, visual and voice identification of a rape suspect under different levels of illumination. Journal of Applied Psychology, 71, 363-370.

Yarmey, A.D. (1993). Adult age and gender differences in eywitness recall in eyewitness recall in field settings. Journal of Applied Social Psychology, 23, 1921-1932.

Yarmey, A.D., & Jones, H.P.T. (1983). Is the psychology of eyewitness identification a matter of common sense?. In S. Lloyd-Bostock & B.R. Clifford (Eds), Evaluating Eyewitness Evidence, pp. 13-40 Chichester: John Wiley & Sons.

Yarmey, A.D., Yarmey, A.L., & Yarmey, M.J. (1994). Face and voice identifications in showups and lineups. Applied Cognitive Psychology, 8, 453-464.

Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus and rapidity of habit-formation. Journal of Comparative and Neurological Psychology, 18, 459-482.

Young, P.C. (1925). An experimental study of mental and physical functions in the normal and hypnotic states. American Journal of Psychology, 36, 214-232.

Yuille, J.C. (1984). Research and teaching with police: a Canadian example. International Review of Applied Psychology, 33, 5-23.

Yuille, J.C. (1993). We must study forensic eyewitnesses to know about them. American Psychologist, 48, 572-573.

Yuille, J.C., & Cutshall, J.L. (1986). A case study of eyewitness memory of a crime. Journal of Applied Psychology, 71, 291-301.

Yuille, J.C., Davies, G., Gibling, F., Marxen, D., & Porter, S. (1994). Eyewitness memory of police trainees for realistic role plays. Journal of Applied Psychology, 79, 931-936.

Yuille, J.C., & Kim, C.K. (1987). A field study of the forensic use of hypnosis. Canadian Journal of Behavioural Science, 19, 418-429.

Yuille, J.C., & McEwan, N.H. (1985). Use of hypnosis as an aid to eyewitness memory. Journal of Applied Psychology, 70, 389-400.

Zelig, M., & Beidleman, W.B. (1981). Investigative hypnosis: A word of caution. Internation Journal of Clinical and Experimental Hypnosis, 29, 401-412.

APPENDICES

Note:-

The questionnaires shown in the following Appendices have been reduced in size from those used in the Experiments.

APPENDIX 8.2: QUESTIONNAIRE 1

Questionnaire

Please answer the following questions as accurately as is possible. When this questionnaire has been completed the experiment will have finished.

1. What is your name? ______________________________

2. What is your age? _____ years _____ months

3 What is your sex? (Tick appropriate box.)

Male.....[ ]
Female...[ ]

4. The fifty photographs, from which you chose twenty, are presented on the sheet in front of you. Please attempt to identify as many of the twenty photographs that you chose as you can. Note down the letter and number of each photograph identified in the space provided below. Do not worry if you are unable to remember every photograph.

1).....	11).....
2).....	12).....
3).....	13).....
4).....	14).....
5).....	15).....
6).....	16).....
7).....	17).....
8).....	18).....
9).....	19).....
10).....	20).....

Turn over.

A number of the following questions will ask you if you agree or disagree with a statement. You will be given a range of answers, place a tick in the box which corresponds most accurately to your opinion of the statement. There will be questions asking about your confidence in various answers that you have provided. Again, a range of answers will be provided, tick the box which most closely corresponds to your confidence in your answer.

5. Do you agree with the statement:
"It was more easy to identify males than females."

strongly disagree	disagree	slightly disagree	no difference	slightly agree	agree	strongly agree

6. How confident are you that your answer to question 5 is accurate?

not at all confident		not very confident		confident		very confident

7. Do you agree with the statement:
"It was easier to identify old people than it was to identify young people."

strongly disagree	disagree	slightly disagree	no difference	slightly agree	agree	strongly agree

8. How confident are you that your answer to question 7 is correct?

not at all confident		not very confident		confident		very confident

9. In the twenty photographs that you chose were there males and females. Please tick the appropriate box if you thought there were (i) more males than females, (ii) more females than males, (iii) equal numbers of males and females.

More males than females.................[ ]
More females than males.................[ ]
Equal numbers of males and females......[ ]

10. How confident are you in your answer to question 9 ?

not at all confident		not very confident		confident		very confident

11. How many photographs do you think that you accurately identified? (Place number on line below.)

______

12. How confident are you that your answer to question 11 is correct?

not at all confident		not very confident		confident		very confident

Thank you for your co-operation in this experiment.

Questionnaire

Please answer the following questions as accurately as is possible.

1. What is your name? ______________________________

2. What is your age in years? _____

3 What is your sex? (Tick appropriate box.)

Male.....[ ]
Female...[ ]

A number of the following questions will ask you if you agree or disagree with a statement. You will be given a range of answers, place a tick in the box which corresponds most accurately to your opinion of the statement. There will also be questions asking about your confidence in various answers that you have provided. Again, a range of possible answers will be provided, tick the box which most closely corresponds to your confidence in your answer.

4. How many photographs do you think that you could accurately identify? (Place number on the line below)

____

5. How confident are you that your answer to question 4 is correct?

not at all confident	not very confident		confident		very confident	

6. Do you agree with the statement:
"It was easier to identify males than females."

strongly disagree	disagree	slightly disagree	no differ-ence	slightly agree	agree	strongly agree

7. How confident are you that your answer to question 6 is accurate?

not at all confident	not very confident		confident		very confident	

8. Do you agree with the statement:
"It was easier to identify old people than it was to identify young people."

strongly disagree	disagree	slightly disagree	no difference	slightly agree	agree	strongly agree

9. How confident are you that your answer to question 8 is correct?

not at all confident		not very confident		confident		very confident

10. In the twenty-five photographs that you chose, there were both males and females. Please tick the appropriate box if you thought that,in your sample, there were (i) more males than females, (ii) more females than males, (iii) equal numbers of males and females.

More males than females.................[ ]
More females than males.................[ ]
Equal numbers of males and females......[ ]

11. How confident are you that your answer to question 10 is correct?

not at all confident		not very confident		confident		very confident

12. If you were asked to testify in court, how many of the twenty-five photographs that you saw, would you be absolutely certain that you saw before?

______

APPENDIX 9.2.2: ANSWER SHEET 1

1. What is your name? ____________________________

2. What is your age? _____ years _____ months

3. What is your sex? (Tick appropriate box.)

4. The fifty photographs from which you chose twenty-five, are presented on the sheet in front of you. Please attempt to identify as many of the twenty-five photographs that you chose as you can. Note down the letter and number of each photograph identified in the spaces provided below. Do not worry if you are unable to remember every photograph.

....

....

....

....

....

APPENDIX 14.2: QUESTIONNAIRE 2

Please fill in this questionnaire.

1. What is your name? ____________________________

2. What is your age? ____ years

3 What is your sex? (Tick appropriate box.)

Male.....[ ]
Female...[ ]

You will be asked some questions about the film which you have just seen. After each question you will be given a choice of two alternative answers. Tick the box which you think is the correct answer. *You must make a choice*, even if it is only a guess. After you have made your choice you will be asked to rate how confident you are in your answer. To do this, circle the number that corresponds with how confident you are.

1. How many people were shown in the film clip?

Four [ ]
Five [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

2. What was the name of the school?

The Willowby Day Centre [ ]
The Willoughby Day Care Centre [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

3. What sex was the person in the bed?

Male [ ]
Female [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

4. How many medicine bottles were on the bedside cabinet?

eight [ ]
nine [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

5. What colour was the hair of the person who was in the bed?

Grey [ ]
Black [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

6. In which hand did the doctor hold his bag?

The left hand [ ]
The right hand [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

7. Did the man who sat on the bed have glasses?

Yes [ ]
No [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

8. When the doctor was writing the prescription on the chest of drawers there were some ornaments on the chest. What animals were they?

cows [ ]
sheep [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

9. Was the man who sat on the bed old or young?

Young [ ]
Old [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

10. Was there a book in front of the photograph on the dresser?

Yes [ ]
No [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

11. Was the person in the bed old or young?

Young [ ]
Old [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

12. How many pens did the man who opened the door to the child have in his pocket?

One [ ]
Two [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

13. What was the present, given to the person in the bed, from the children?

A picture they had drawn [ ]
A stuffed toy that they had made [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

14. How many times did the child knock on the door?

14 [ ]
15 [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

15. Who knocked on the front door?

A young boy [ ]
A young girl [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

16. What was the number on the child's sleeve?

416 [ ]
418 [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

17. What did the child cuddle?

A rag doll [ ]
A puppy dog [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

18. How many panes of glass were in the door which the child knocked on?

eight [ ]
ten [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

19. What sex was the doctor?

Female [ ]
Male [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

20. Was there a picture of a zebra or a horse on the wall?

Yes [ ]
No [ ]

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

APPENDIX 15.2.2: QUESTIONNAIRE 3

Please fill in this questionnaire.

1. What is your name? ______________________________

2. What is your age? ____ years

3 What is your sex? (Tick appropriate box.)

Male.....[ ]
Female...[ ]

You will be asked some questions about the film which you have just seen. After each question there will be a space for you to provide an answer. You *must* answer each and every question. If you do not know the answer you must still make a response even if it is only a guess. After you have made your choice you will be asked to rate how confident you are in your answer. To do this, circle the number that corresponds with how confident you are.

1. What implement was being washed in the sink?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

2. What was behind the Tabasco sauce bottle in the kitchen?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

3. How many times did the woman wind up the clockwork toy before she put it on the kitchen counter?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

4. What did the woman do after eating the fruit?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

5. How many bottles were on top of the bathroom cabinet?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

6. What was the dialling code of the first phone number on the board at the start of the film?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

7. Who was the poet mentioned on the radio?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

8. What was the vegetable which the woman was cutting by the cooker?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

9. What time did the clock say?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

10. What song was the woman singing?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

11. How may tiles were above the bathroom sink?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

12. What fruit did the woman eat?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

13. What was the name of the person referred to in the woman's song?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

14. What was the brand name of the washing up liquid?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

15. What clockwork toy was shown in the film?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

16. What was the man on the television programme drinking?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

17. What kind of hat was the dead man wearing?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

18. Which cupboard had a broken handle?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

19. What was the soap in the bathroom shaped to look like?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

20. What was the man shown doing in the bathroom at the beginning of the film?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

21. What was the title of the big book?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

22. How many pieces did the woman cut the fruit into?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

23. With what was the bathroom wall decorated?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

24. Where was the telephone situated?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

25. What card was lying against the wine bottle?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

26. A picture of what animal was on the bathroom door?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

27. What kind of clock was shown in the film?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

28. How many characters, apart from the cameraperson, were shown in the film?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

29. What was the cuddly toy on the table?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

30. What was hanging on the wall to the right of the kitchen window?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

31. Where was the dead body found?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident		fairly confident			very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

32. What brand were the tinned tomatoes on the kitchen counter?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

33. What was on the dish next to the television set?

____

Circle the number which most closely matches how confident you are that your answer to this question is correct.

pure guess		slightly confident			fairly confident		very confident		absolutely sure
1	2	3	4	5	6	7	8	9	10

APPENDIX 17.2.2: ANSWER SHEET 1B

What is your name? ____________________

All fifty photographs, from which you chose twenty-five, are presented on the identification sheet in front of you. Please attempt to identify as many of your twenty-five photographs as is possible. Note down the letter and number of each photograph identified in the spaces provided below. Please note how confident you are in each identification you make by ticking the appropriate box.

Photograph identified (e.g. G1)	How confident are you in your identification?				
	not very confident	slightly confident	confident	very confident	absolutely certain

APPENDIX 22.1

THE INFLUENCE OF ITEM DIFFICULTY ON THE RELATIONSHIP BETWEEN EYEWITNESS CONFIDENCE AND ACCURACY.

Kebbell, M.R., Wagstaff, G.F. & Covey, J.A. (In press.) British Journal of Psychology.

Abstract

Research indicates that the confidence which eyewitnesses express in information heavily influences both the investigative process, and the credence which jurors give to eyewitness testimony. However, studies in this area suggest that there is either no relationship, or only a small positive relationship between eyewitnesses' confidence and accuracy. Nevertheless, it is argued here that researchers may have paid insufficient attention to the issue of item difficulty, and have used statistical procedures that fail to consider highly accurate responses with low variance. In an attempt to address these issues, two experiments were conducted which measured C-A relationships in response to information seen in video films. In each case, questions were used that ranged in difficulty. Higher C-A correlations than are usually reported were found in both experiments. Furthermore, when subjects were 'absolutely certain' that a piece of information was correct they almost invariably were accurate. Possible practical implications of these findings are discussed.

Introduction

Research has shown that the confidence which eyewitnesses express in information heavily influences both the investigative process and the credence which jurors give to eyewitness testimony; if eyewitnesses express certainty that their answers are correct, their responses are more likely to be perceived as correct (Brigham & Wolfskeil, 1983; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Thomas, 1988; Fox & Walters, 1986; Leippe, Manion & Romanczyk, 1992; Wells, Ferguson & Lindsay, 1981; Lindsay, Wells & O'Connor, 1989).

However, in contrast with the intuitions of investigators and jurors that there is a strong positive relationship between eyewitnesses' confidence and their accuracy, much research in this area appears to contradict this assumption. Reviews by, for example, Bothwell, Deffenbacher, & Brigham (1987), Deffenbacher (1980), Fruzzetti, Tolland, Teller & Loftus (1992) and Wells & Murray (1984), suggest that there is either no relationship, or only a small positive relationship between eyewitnesses' confidence and their accuracy.

In an attempt to explain these apparently counter-intuitive findings, Smith, Kassin & Ellsworth (1989) have suggested that researchers have concentrated on the confidence-accuracy (C-A) relationships 'between-subjects', comparing the accuracy of confident witnesses to less confident witnesses, rather than the relationship within subjects' own statements. In the latter case, an eyewitness may say that he/she is absolutely certain of some things but is not at all certain of others. Indeed, what Smith et al. term 'within-subject' C-A relationships are equally, if not more likely than between-subject C-A relationships to have practical implications in forensic situations for both investigators and jurors.

To assess within-subject and between-subject confidence accuracy relationships Smith et al.(1989) showed subjects a slide presentation followed by a number of two-alternative forced choice questions. They were then required to rate their confidence in each answer on a ten-point scale. The average between-subjects and within-subjects C-A correlations were comparatively low; $r = .14$ and $r = .17$ for between and within subjects measures respectively. Smith et al. concluded: 'Confidence is not a good predictor of accuracy. Common sense and the Supreme Court notwithstanding, confidence is not a useful indicator of the accuracy of a particular witness or of the accuracy of particular statements made by the same witness' (p.358).

However, Perfect, Watson & Wagstaff (1993) note that Smith et al. assessed memory with a forced two-choice recognition procedure that gave a hit rate of only 63%. As a large number of these hits (37%) would have occurred by chance, they suggest that this high guessing rate may be in part responsible for the low correlations found. Perfect et al. (1993) therefore conducted a further investigation into within subject and between subject C-A correlations in eyewitness performance. Subjects in the eyewitness condition viewed a short film clip. To reduce the chances of producing correct answers by guessing, subjects were required to answer 35, five-alternative forced-choice questions and to rate their confidence in each on a five point scale ranging from 'very confident' to 'no idea'. They found an higher overall correlation than Smith et al. for the between-subjects analysis (Goodman-Kruskal Gamma = .49), but no correlation for the within-subjects analysis (Gamma= -.03). So currently the experimental literature still suggests that there is little consistent relationship between eyewitnesses confidence and accuracy, even if 'within-subjects' relationships are considered.

Howver, one factor that has yet to be systematically investigated is item difficulty. Typically in work in this area, researchers attempt to select items so as to avoid floor and ceiling effects; i.e. they try to avoid items that are either very easy or very hard to remember. But in real-life forensic investigations some questions that eyewitnesses are asked may be easier to answer than others. For example, in the case of an assault, an eyewitnesses may be asked 'what sex was the attacker?' Gender is amongst one of the first item noticed about an individual and is very likely to be answered accurately. Furthermore, most eyewitnesses are likely to be very confident that the identification of an individual's sex is correct. Alternatively, if eyewitnesses are asked, for example, 'what was the colour of the attacker's eyes?', this question might be more difficult to answer and eyewitnesses may be less confident about their accuracy (see Christianson & Hubinette, 1993).

It may be the case, therefore, that in an attempt to avoid ceiling and floor effects, previous researchers may have chosen unrealistic and overly homogeneous pools of items, thus reducing the variance necessary for high correlations. Consequently, higher C-A relationships might result if a heterogeneous range of 'hard' or 'easy' questions were used.
Lower C-A correlations would be expected, however, if the categories of items were considered separately.

Another possibly important related factor is that of the relationship between 'absolutely certain' responses and accuracy. This effect may be precluded when 'easy' items are excluded, and, as it is not necessarily related to correlation size, it may often missed in correlational analysis (Gruneberg and Sykes, 1993). Eyewitnesses are frequently asked by the Police and Lawyers if they are 'absolutely certain' about information that they provide so that the Police and Jurors can evaluate the accuracy of information that they are given.

Regardless of overall C-A accuracy, it could be the case that the relationship between these 'absolutely certain' responses and accuracy remains high.

The purpose of the following two experiments was to investigate these issues.

EXPERIMENT 1

Introduction

The first experiment was constructed to be similar to that of Smith et al. (1989). Subjects were required to answer two-alternative forced-choice questions about a film which they had watched. However, it differed from the experiment of Smith et al. in that questions were constructed to fall into one of two categories, easy or hard.

Method

Subjects

Subjects were 51 prospective University students (37 females, 14 males) visiting Liverpool University Psychology Department for an open day (mean age= 19 years, range 17-29, SD=3.40).

Apparatus and Procedure

Subjects were tested in groups varying in size from 14 to 21. They were shown a five and a half minute colour film via a video player and television monitor. The film concerned an elderly couple and their doctor in their home.

Following the film subjects were given a five minute filler task, followed by a questionnaire devised to test their recall of the film. The format was similar to that used by Smith et al. (1989). Twenty questions were used, each followed by a forced choice of two alternative answers. The questions were devised by two experimenters so that there were ten questions in each of two categories; easy or hard. For example, an easy question was 'What sex was the person in the bed?', while a hard questions was, 'was there a picture of a zebra or a horse on the wall?' Respondents were required to rate their confidence in each answer on a ten point Likert scale ranging from 'pure guess' (1) to 'absolutely certain'(10).

Results and discussion

Subjects were significantly more likely to answer easy questions correctly than hard questions, $t(50)=20.22$, $p<.0001$, thus confirming the experimenters' categorisation of question difficulty. The numbers of accurate and inaccurate responses for easy and hard questions are shown in Table 1.

A C-A correlation was calculated for each subject across the 20 questions (easy and hard questions combined). This resulted in 50 of what Smith et al. (1989) refer to as 'within subjects' correlations (the *N*'s for various calculations vary because in some cases, when all questions were answered correctly or incorrectly it was not possible to use the data). These correlations were then averaged. The average within-subject correlation was $r=.54$, SD= 0.15. To test whether this correlation was significantly different from zero, the procedure used by Smith et al. (1989) was adopted. The the correlation coefficient for each subject was transformed into a z score; the z scores were averaged and this average was tested against zero. The result was significant, $t(49)=3.18$, $p<.005$.

Within-subject C-A relationships were also calculated for easy and hard questions independently; i.e. each subject's C-A correlation was calculated for the 10 easy or 10 hard questions then averaged across subjects. These correlations were tested against zero again using converted z scores, but neither was significant. Taken together, as predicted, these results suggest that using questions of varied difficulty, *within-subjects*, may produce higher C-A correlations than using questions of a similar difficulty.

The C-A correlation across-subjects for each of the 20 questions (easy and hard combined) was calculated; i.e. what Smith et al., term 'between subjects' correlations. This resulted in 16 correlations. The average of these correlations was not significantly different from zero. The between subjects C-A correlations was calculated for the six easy questions, and 10 hard questions. Neither was significantly different from zero. On the whole, therefore, the 'between subjects' C-A relationships seem to be low and in line with previous findings.

Further, each subject's average accuracy was correlated with his/her average confidence rating. Overall (easy and hard questions combined) this correlation was $r=.26$, $p<.10$ (N= 51). However, whereas for hard questions alone this correlation was only $r=.21$, $p>.10$, for easy questions it was $r=.54$, $p<.0001$. Thus more confident subjects were only more accurate than less confident subjects on easy questions.

In addition, the average accuracy rate and average confidence score for each question was correlated. For all 20 questions (easy and hard combined) this correlation was $r=.74$, $p<.0005$. For the ten easy questions the correlation was significant $r=.74$, $p<.05$; but not for the ten hard questions, $r=.12$, $p>.10$. Thus, overall, the greater the confidence that subjects expressed in answers to a particular question the greater the likelihood that subjects would answer that question correctly. This relationship was maintained when easy questions, but not hard questions, were considered independently.

Subjects' average confidence scores were then calculated for the categories of easy questions answered correctly, easy questions answered incorrectly, hard questions answered correctly and hard questions answered incorrectly. These are summarised in Table 1. A two-way repeated measures ANOVA (2 X 2, question difficulty X correct/incorrect answer) was conducted on these data (only 12 of the 51 subjects could be used for this analysis as 36 subjects answered all of the easy questions correctly, two subjects answered all of the easy questions correctly and all of the hard questions incorrectly, and one subject answered all of the easy and the hard questions correctly).

Results showed that average confidence was significantly higher in correctly answered questions than in incorrectly answered questions, $F(1,11)=43.25$, $p<.0001$, and the average confidence expressed in easy questions was significantly greater than that expressed in hard questions, $F(1,11)=167.81$, $p<.0001$. A significant interaction was also found between question difficulty and correct/incorrect answers $F=(1,11)=10.84$, $p<.01$. Follow up univariate F tests showed that all the means were significantly different ($p<.05$) from each other, but whilst the difference between average confidence in correct easy questions, was very much greater than the average confidence expressed in easy incorrect answers, the difference between average confidence in hard questions which were answered correctly and incorrectly, although in the same direction, was not so great.

It can also be noted here that subjects reported greater confidence in their correct responses to easy questions than to hard questions, and that this relationship was maintained for incorrect answers; i.e. subjects reported greater confidence in easy questions which were answered incorrectly than hard questions which were answered incorrectly.

\-----------------------

insert table 1 about here

\-----------------------

Out of a total of 1016 overall responses, 321 of the answers given by subjects were rated as being 'absolutely certain' that their answers were correct. Of these 321 answers 319 were correct, an accuracy rate of 99.4%; thus when subjects were 'absolutely certain' in a response then they were unlikely to be inaccurate.

Although the manipulations used in this study were relatively successful in raising 'within subjects' C-A relationships, it could still be argued that, as in the experiment by Smith et al. (1989), because many questions could be answered correctly by chance alone, the C-A relationships were artificially lowered. To control for this possibility, a second experiment was conducted.

EXPERIMENT 2

Introduction

In experiment 2, to reduce the effects of guessing, rather than using two-alternative forced choice questions, open-ended questions were used. Also, as a simple selection of 'easy' and 'hard' questions might be considered to be over restrictive and unrealistic, questions were selected so as to be divided into three categories of difficulty: easy, medium, and hard.

Method

Subjects

Subjects were 45 undergraduate Psychology students (32 females, 13 males). The mean age was 23 years (range 18-43, SD= 6.10).

Apparatus and Procedure

Subjects were tested in two groups, one of 17 and one of 28. They were shown a 5 and a half minute black and white video film that concerned the implied murder of a male by a female.

Following the film, subjects were given a 10 minute filler task, followed by a 33 item questionnaire devised to test their recall of the film. The format of the questionnaire was similar to that of experiment 1. The questions were open-ended, but subjects were required to provide an answer, even if this was only a guess. The questionnaire was devised by two experimenters who agreed on 11 questions in each of three categories of item difficulty; easy, medium or hard. To reduce the influence of being correct by chance, each question was devised such that a range of plausible answers was possible. For example, an easy question was 'what song was the woman singing?', a medium difficulty question was 'what was on the dish next to the television set?', and a hard question was 'what was behind the Tabasco sauce bottle in the kitchen?'

After answering each question subjects were required to rate their confidence in their answer on a ten point Likert scale as for experiment 1.

Results

A one-way repeated-measures ANOVA showed a significant effect of question difficulty on the number of correct answers $F(2,88)=591.37$, $p<.0001$. Follow up F tests ($p<0.05$) confirmed that these differences were in the appropriate direction; easy questions were more likely to be answered correctly than medium questions, which in turn were more likely to be answered correctly than hard questions. This verified the experimenters' categorisation of item difficulty (see Table 2).

A C-A correlation was calculated for each subject across the 33 questions and these correlations were averaged for the 45 subjects. The average within-subject correlation (easy, medium and hard combined) was $r=.78$, SD=.08, which was significantly different from zero, $t(44)=5.00$, $p<.005$.

Within-subject C-A relationships were also calculated for easy, medium and hard questions independently; i.e. each subject's C-A correlation was calculated for the 11 easy, medium or hard questions then averaged across subjects. For easy questions the average within-subjects correlation was $r=.76$, SD=.21, which was significantly different from zero, $t(23)=2.261.04$, $p<.05$ (only 24 of the 45 subjects were used in this analysis as 21 subjects answered all of the easy questions correctly). For the medium difficulty questions the average within-subjects correlation was $r=.55$, SD=0.23, which was also significantly different from zero, $t(44)=1.82$, $p<.05$. However, for the hard questions the average within-subjects correlation (r= ?) was not significantly different from zero, $t(19)=0.73$ (only 20 of the 45 subjects could be used in this analysis as 25 subjects answered all of the hard questions incorrectly).

Again, as predicted, on the whole these results suggest that using questions of varied difficulty, *within-subjects*, may produce higher C-A correlations than are usually found with questions of a similar difficulty. However, this relationship also remained when easy and medium difficulty questions were considered independently.

The C-A correlation across-subjects for each of the 33 questions (easy, medium and hard combined) was then calculated, and these 'between-subject' correlations were averaged. The average between-subjects correlation was

r=.49 (SD=.35), which was not significantly different from zero, $t(25)=1.20$ (only 26 of the 33 questions could be used in this analysis as three of the easy questions were answered correctly by all subjects and four of the hard questions were answered incorrectly by all subjects). Between-subject C-A performance was then separately calculated for easy, medium and hard questions. These three correlations were not significantly different from zero. Thus these findings show little support for a between-subject C-A correlation.

Each subject's average accuracy was correlated with his/her average confidence rating (N= 45). Overall (easy, medium and hard questions combined) the correlation was r=.56, p<.0001. For easy questions alone this correlation was r=.69, p<.0001, for medium questions r=.45, p<.01, and for hard questions r=.45, p<.01. These results suggest that in this experiment, overall, more confident subjects were also likely to be more accurate.

In addition, the average accuracy rate and average confidence score for each question was correlated (between-subjects). For all 23 questions (easy, medium and hard combined) this correlation was r=.97, p<.0001. For the 11 easy questions the correlation was r=.96, p<.0001; for the 11 medium difficulty questions, r=.84, p<.005 and for the 11 hard questions, r=.29. Again, this shows that, overall, the greater the confidence that subjects expressed in answers to a particular question the greater the likelihood that subjects would answer that question accurately. This relationship was maintained when easy questions and medium difficulty questions were considered independently, but not for hard questions.

Subject's average confidence scores were also calculated for six categories of question response: easy questions answered correctly; easy questions answered incorrectly; medium questions answered correctly; medium questions answered incorrectly; hard questions answered correctly; and hard questions answered incorrectly. These data are summarised in Table 2. A two-way repeated-measures ANOVA (3 X 2, question difficulty X correct\incorrect answer) was conducted on these data (only 12 of the 45 subjects could be used for this analysis because many subjects answered all of the easy questions correctly, all of the hard questions incorrectly or both). There was a significant main effect for question difficulty, $F(2,20)=26.53$, p<.0001; follow up univariate F tests showed that all means were significantly different from each other. Subjects expressed greater confidence the easier the questions. There was also a significant main effect for confidence in correct/incorrect answers. Average confidence was higher for correctly answered questions than for incorrectly answered questions, $F(1,10)=96.18$, p<.0001. The interaction was not significant.

----
insert table 2 about here

Out of a total of 1481 overall responses, 387 of the answers given by subjects were rated as being 'absolutely certain' that their answers were correct. Of these 387 answers 377 were correct, an accuracy rate of 97.4%, again supporting the assertion that subjects who rate their confidence as 'absolutely certain' are very unlikely to be inaccurate.

General Discussion

The results of both studies show that when questions which varied in difficulty are used, within-subjects confidence accuracy relationships are higher than have been previously been reported. Indeed, subjects appear quite able to accurately calibrate their accuracy in answering a question.

The high overall within-subjects C-A correlations were likely to have been facilitated by the presence of 'easy' and 'hard' questions. Subjects were often sure that they 'knew' the correct answers to the easy questions. Therefore, they responded that they were 'absolutely certain' that their answers to these questions were correct. These answers were indeed almost invariably correct. For the 'hard' questions subjects were often sure that they did not 'know' the answer and that they were guessing so they made 'pure guess' responses concerning the likelihood of these answers being accurate. These answers were rarely correct.

Previous studies, which ignored easy and hard questions may have had few 'absolutely certain' or 'pure guess' responses due to a lack of easy or hard questions. Eyewitnesses may be able accurately identify questions to which they *definitely* can or *definitely* cannot accurately and express appropriate confidence judgements in these answers. However, they may be less able to gauge their own confidence, and also less able to articulate this confidence, in the accuracy of answers which are between these two extremes of difficulty.

Within-subject C-A relationships were calculated independently for different levels of question difficulty. No significant correlations were found in experiment one. This may have been influenced by two factors. For the easy questions the number of subjects which could be used for correlational analysis was low as a number of subjects accurately answered the all questions correctly. In the hard group the two-alternative forced choice nature of experiment 1 meant that through chance 50% of guesses would be correct (easy questions were not affected to this extent due to the high accuracy rate- subjects would not have guessed so much because they 'knew' the answers). Thus, correct answers which were in fact guesses (with accordingly low confidence ratings), would affect the C-A relationships.

In experiment two there were significant within-subject C-A correlations for questions of easy and medium difficulty but not for hard questions. The reason for this discrepancy between easy and medium difficulty questions and hard questions may be a difference in the kind of errors which occurred for different question types. For easy questions the majority of subjects provided correct answers and displayed high levels of confidence in the accuracy of those answers. What few errors did occur may have been due to lapses of attention, for example by subjects being momentarily distracted and therefore missing target information. Those subjects who missed some 'easy' stimuli would express low confidence and show poor accuracy to questions concerning information that they had missed. This combination of high confidence and accuracy for most easy questions with a few questions which were answered with low confidence and accuracy because of 'missed' information would produce a high C-A correlation for easy questions.

However, for hard questions even subjects who paid close attention to the film may have been unable to answer the questions correctly and so expressed low confidence in the accuracy of their responses combined with a low accuracy rate.

Of the few accurate responses many could have occurred through guessing. These would be indistinguishable from inaccurate 'pure guess' responses in terms of confidence ratings, thus decreasing any C-A relationship. Subjects who were distracted in this situation would therefore be indistinguishable from those who answered incorrectly but attended to the stimuli. Further, when considering the finding that the C-A relationship for hard questions is not significantly different from zero, it must be remembered that the original hypothesis suggested that in such situations of similar question difficulty such an effect may occur.

The repeated-measure ANOVA tests (in effect a within-subjects measure of performance) considering average confidence ratings with respect to question difficulty and accuracy for both experiments can further be taken to support a strong C-A relationship. Both experiments showed that confidence was significantly greater for correct answers than incorrect answers.

Furthermore, another apparent effect of question difficulty was revealed. Both experiments showed question difficulty to effect subject's confidence ratings of correct and incorrect answers when they were considered independently. For example, for correct answers to experiment 2 with subjects expressed more confidence in easy questions than in questions of medium difficulty. In turn greater confidence was expressed in questions of medium difficulty than in questions which were hard. A similar pattern emerged for experiment 1 for correct answers and in both experiments for incorrect answers.

This suggests that subjects may not only use an internal 'feeling of knowing' of whether their answer is correct to determine their confidence rating but also consider additional components. One such factor may be an individual's evaluation of how confident that they '*should*' be. This may take into account how able they should be given the circumstances in which they witnessed the event, i.e. is this something which I think I should be able to remember?

When subject's average accuracy was correlated with their average confidence rating significant correlations were found in the second experiment both overall and when each of the three types of question difficulty which were considered. In the first experiment there was an overall trend for the two factors to be correlated but this was only significant for easy questions. The first experiment may have been affected by the high level of accurate answers produced by chance in the hard question group. For the hard questions in experiment 1 through chance some (50%) of guesses would be correct (easy questions were not affected to this extent due to the high accuracy rate and therefore a low guessing rate). Thus, correct answers which were in fact guesses (with accordingly low confidence ratings), would affect measures which were made. This suggests that subjects which express greater confidence overall may be more likely to be accurate.

In addition, the average accuracy rate and average confidence score for each question was correlated. In the first experiment this was significant overall and for easy questions. Again there was no significant correlation for hard questions, which as mentioned previously may be influenced by the high number of accurate answers produced by chance in this group. In the second experiment there was a significant effect overall and for questions which were categorised as being of easy or medium difficulty but not for hard questions. The overall effects are not surprising, easy questions have a high accuracy rate and high confidence is expressed in answers, while the reverse is true of hard questions, with medium questions somewhere in between. The variation in the average confidence scores and the average accuracy scores range from very high confidence and accuracy for easy questions while the hard questions show low confidence responses and low accuracy. So it is not surprising that there is a strong overall correlation. The smaller correlations found when questions of each difficulty are considered independently is likely to also be due to such variations albeit on a smaller scale.

However, when between-subjects C-A performance was considered the two experiments found small, positive but non-significant C-A relationships, irrespective of whether overall performance or performance on easy, medium or hard questions were considered independently. This may be due to differences in the confidence of individual subjects, i.e. some subjects may be generally more confident than others. In within-subjects C-A correlations such subjects are compared against their own relative confidence judgements while in between-subject situations they are compared to one-another. Although a relationship was found between average confidence and average accuracy this appears not to be strong enough to produce a between-subjects effect.

The above experiments raise the issue of what is the most appropriate paradigm to use in investigating C-A relationships for eyewitness situations i.e. should the questions which are used be of similar or varying difficulty? Using questions of varying utility may be an appropriate way to investigate eyewitness C-A relationships. The utility of information in 'solving' crimes, forensic utility, is not necessarily proportionally related to item difficulty. The idea that questions must be difficult in order to obtain useful information may be misleading. For example, establishing the gender of a suspect is relatively easy, yet it is of considerable use to an investigation, immediately eliminating half the population from future enquiries. Thus, information gathered from easy questions may be very relevant to eyewitnesses' testimony. Newlands and George (1994) have investigated which factors Police officers rate as being most useful to the successful conclusion of a Police investigation. Perhaps future work could use these factors to investigate whether a strong confidence-accuracy correlation remains for questions that are forensically useful.

In many respects the information reported here represents the best C-A relationship which is likely to be found. Other variables in 'real-life' situations may reduce this relationship. For example, in 'real-life' situations eyewitnesses are not required to rate their own confidence on a Likert scale, instead their confidence is gauged by investigators or jurors through a variety of other factors such as tone of voice, eye contact as well as requesting eyewitnesses to express how sure they are of an answer. Further factors, such as social pressure in interview or trial situations may make eyewitnesses less or more confident in parts their statements (depending on what kind of pressure is applied) or longer retention intervals may change the C-A relationship.

Future work may also wish to investigate the influence of interview technique on C-A relationships. For example 'hypnosis' has been shown in some situations to increase subject's confidence in inaccurate information. Also, the Cognitive Interview technique as described by Fisher and Geiselman (1992, p.38) contains instructions for interviewers to increase eyewitnesses confidence. It is unclear what effects these interview techniques have on C-A relationship.

In sum, these experiments support the conclusion that eyewitnesses' judgements of confidence, within their own testimony, may be usefully considered in judging the likelihood of accuracy, especially if they are 'absolutely certain' that an answer is correct. However, there are a number of other factors which may require consideration in 'real-life' situations that are likely to mediate this relationship.

Acknowledgements

This research was supported by a grant to the first two authors from the Economic and Social Research Council (grant no. R00429234159). We would also like to acknowledge the contribution of Andy Taylor, Beth Cooper and Paddy Paterson in the production of the video film for experiment 2 and of Mark Craigie for statistical advice.

Table 1. Scores in relation to confidence ratings, difficulty of question and accuracy.

Measure	Easy questions	Hard questions	Overall
Average No. correct	9.7 SD=.72	5.2 SD=1.5	15 SD=1.8
Average No. incorrect	.33 SD=0.71	4.8 SD=1.5	5.1 SD=1.9
Average confidence rating of correct answers	9.0 SD=0.73 (8.7 SD=0.76)	2.7 SD=1.3 (3.3 SD=1.8)	6.95 SD=.80 (6.0 SD=1.0)
Average confidence rating of incorrect answers	4.9 SD=2.4 (4.9 SD=2.4)	2.1 SD=1.2 (2.0 SD=1.1)	2.2 SD=1.3 (3.4 SD=1.7)

Note. Figures in brackets were used for the analysis of variance. These are different because a repeated-measures ANOVA was used, consequently some subjects had to be dropped as they either answered all easy questions correctly or all hard questions incorrectly or a combination of these.

Table 2. Breakdown of scores with respect to confidence ratings, difficulty of question and accuracy.

Measure	Item difficulty			
	Easy	Medium	Hard	Overall
Average No. correct	9.9 SD=1.2	4.6 SD=2.0	.67 SD=.80	.15 SD=2.8
Average No. incorrect	1.1 SD=1.2	6.4 SD=2.0	10 SD=.82	18 SD=2.8
Average confidence rating of correct answers	9.00 SD=.90 (9.18 SD=.74)	6.3 SD=2.1 (7.06 SD=1.3)	3.4 SD=3.0 (3.88 SD=3.3)	7.9 SD=.98 (6.7 SD=1.2)
Average confidence rating of incorrect answers	3.02 SD=2.43 (4.21 SD=2.9)	2.63 SD=1.13 (3.30 SD=1.6)	1.41 SD=0.48 (1.37 SD=.38)	1.93 SD=0.62 (3.0 SD=1.1)

Note. Figures in brackets were used for the analysis of variance. These are different as a repeated-measures ANOVA was used some subjects were dropped as they either answered all easy correctly or hard questions incorrectly.

APPENDIX 22.2

THE COGNITIVE INTERVIEW: AN ANALYSIS OF ITS FORENSIC EFFECTIVENESS.

M.R. Kebbell and G.F. Wagstaff (in press). In D.V. Canter (Ed).
Investigative Psychology

The significance of eyewitness testimomy

When investigating criminal acts, the testimony of eyewitnesses is often of crucial importance. If an eyewitness is unable to identify a criminal or remember details of a crime, then the perpetrator may go unpunished. Conversely, if innocent individuals are falsely identified, they may be convicted of crimes which they did not commit. Also the recall of inaccurate information by an eyewitness may mislead the Police, preventing an appropriate investigation of the crime.

The importance of eyewitness testimony in Western criminal justice systems is well illustrated by research conducted by Sanders (1986). When he asked Sheriffs' deputies and detectives in New York the question, "What is the central and most important feature of criminal investigations?", the majority of respondents replied "eyewitnesses". In addition, the report of the Rand Corporation in 1975 found that the major predictor of whether a crime was solved or not was the completeness and accuracy of the eyewitness account. The testimonies of eyewitnesses also take up a considerable amount of Police time. Research in Germany by Herren (1976) indicates that Police officers spend 70 - 80% of their working time interviewing witnesses, victims and suspects.

Given these considerations, not surprisingly, the topic of eyewitness testimony has received considerable attention from psychologists. Much of the research on eyewitness testimony has tended to emphasise the negative aspects of eyewitness performance, and to specify the circumstances in which eyewitnesses are most likely to produce errors (see for example, Loftus, 1979; Lloyd-Bostock and Clifford, 1983). The findings of such research have led researchers such Fisher, Geiselman & Raymond (1987a) to conclude that eyewitness performance is "incomplete, unreliable, partially constructed and malleable during the questioning procedure", and add that "because of the potential inaccuracy in eyewitness reports, strict reliance on eyewitness identification may often lead to false convictions" (p. 401). One major problem has been the failure to find a consistent positive relationship between confidence and accuracy (e.g. Wells & Lieppe, 1981; Deffenbacher, Brown & Sturgil, 1978; Lindsay & Wells, 1980). This assumes considerable importance as a number of researchers have found that people intuitively believe that an eyewitness's confidence is a predictor of his or her recall accuracy (Yarmey & Jones, 1983; Deffenbacher & Loftus, 1982; Lindsay & Ferguson, 1979).

Nevertheless, frequently the Police have little or no physical evidence, and so must rely on the testimony of eyewitnesses. For this reason more recently methods of enhancing eyewitness recall have received considerable attention. But what, ideally, might a technique designed by psychologists be able to provide a forensic interviewer?

There seem to be three major considerations.

1. The technique should reliably enhance accurate eyewitnesses recall. Accurate recall should be increased without corresponding increases in inaccurate information.

2. The technique should produce testimonies in which eyewitness confidence is related to accuracy, such that the more confident an eyewitness is with respect to an item of information, the more likely that item is to be correct. This might aid the forensic interviewer considerably. If the eyewitness confidence is related to accuracy, then the more confident that an eyewitness is about a piece of information, the greater reliance can be placed on this information in the forensic investigation. Similarly, if confidence is related to accuracy, this might aid jurors in evaluating an eyewitness's testimony.

3. The technique should be easy to use so that Police officers may be trained and to use the technique in a relatively short time, and be practically applicable to real life forensic investigations.

The Cognitive Interview

Recently, considerable interest has been shown in a procedure developed by Geiselman and Fisher to enhance eyewitness performance which they term the 'Cognitive Interview' (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissan & Prosk, 1984; Geiselman & Fisher, 1988; Geiselman & Padilla, 1988; Geiselman, Fisher, MacKinnon & Holland, 1985; Geiselman, Fisher, MacKinnon & Holland, 1986; Fisher, Geiselman & Amador, 1989; Fisher, Geiselman & Raymond, 1987a; Fisher, Geiselman, Raymond, Jurkevich &

Warhaftig 1987b). The main aim of the present chapter is to evaluate the efficacy of this procedure in the light of the above criteria.

Geiselman and Fisher, together with various co-workers, have attempted a systematic approach to their work on enhancing eyewitness recall that has involved four main factors.

1. A comprehensive review of the 'cognitive' literature.

Geiselman and Fisher have evaluated the current theoretical and experimental literature with regards to memory in order to determine what could be usefully applied in the forensic situation.

2. A systematic analysis of real life forensic interviews.

They have undertaken an analysis of real-life forensic interviews to look at the way real interviews are conducted and how they might be improved. This is important as there appear to be some fairly obvious problems with standard forensic interviews that may be dealt with without recourse to complex psychological theories.

3. A comprehensive experimental programme.

They have implemented a comprehensive experimental programme in an attempt to evaluate the relevant issues in a systematic and realistic way.

4. Reporting of findings.

They have reported their findings not only to psychologists, but also to those to whom the results are of most concern, Police officers. This has been achieved by publishing the results in Police journals, for example the "Journal of Police Science and Administration", using language which is easily understood by non psychologists. Indeed they report that the Police have been continually involved at all stages and the relevant data have been reported to them.

Theoretical and experimental Background: Context effects

Part of the development of the Cognitive Interview has involved taking the results from 'pure' cognitive psychology and applying them to the field of eyewitness testimony. Although there are a number of cognitive findings that seem relevant to forensic investigations (Bekerian & Dennett, 1993), the area of 'context' effects has been particularly significant in the case of the Cognitive Interview.

Context effects have been evident in the cognitive literature for many years. The basic assumption underlying these effects is that material will be remembered better if the recall or recognition of the material takes place in the same or similar context to that in which the material was learned. The idea is that reinstating elements of the learning context will provide the subject with retrieval cues to enhance memory. One example is an experiment by Godden and Baddeley (1975), which tested the memory capacity of deep-sea divers. Subjects were asked to learn lists of words, either on a beach or under 15 feet of water. Their recall was tested either in the same environment or in the opposite environment. Godden and Baddeley found that environment had no major effect if recall was conducted in the same environment as encoding. However, if the subjects encoded the information in one environment then were tested in the other, recall was dramatically impaired and subjects remembered approximately 40% less information. The results of Godden & Baddeley have been replicated by other researchers, though with less dramatic effects (e.g. Smith, Glenberg and Bjork, 1978). It could be that the effects are less dramatic because the differences between encoding and retrieval conditions are also less dramatic.

Smith (1979) investigated whether it is necessary to physically reinstate the same environment for context-dependency to work or whether it is sufficient simply to imagine the original environment. He had subjects study words in a distinctive basement room one day, then had them recall the words either in the same room or in a different fifth floor room the next day. Subjects in the basement recalled about 18 words, significantly more than the group tested in the fifth floor room who could only recall about twelve words. A third group was also tested in the room on the fifth floor; however, these subjects were instructed to remember as much as possible of the original learning environment of the room in which they learnt the words before they were required to try and recall the list. This group recalled an average of 17.2 words, which was not significantly different from the average score of those who were tested in the same physical environment, but significantly more than those who were simply tested in the fifth floor room. Hence it appears the original context does not necessarily have to be physically reinstated; simply thinking of the encoding context can enhance recall.

There is also evidence of context effects that are dependent on internal mood states. Goodwin, Powell, Bremer, Hoine & Stern (1969) tested the effects of alcohol on memory tasks. They found similar results to those of

Godden & Baddeley; information encoded when subjects had drunk alcohol was more accurately recalled when subjects had drank alcohol again than when they were sober, and vice versa. Similar findings have been found with happy and sad moods. Teasdale & Fogarty (1979) found that when mood was manipulated subjects who were sad found it easier to recall sad prior experiences than happy experiences.

Thus there seems to be considerable support for context effects, including mental reinstatement, in the experimental literature. Consequently, this idea has been fundamental to the development of the Cognitive Interview.

The standard forensic interview in real life

As well as examining the literature on memory, Fisher and his colleagues conducted an analysis of real life Police interviews with the intention of taking their research "Out of the laboratory and into the field where actual crime interviews are conducted by Police" (Fisher, Geiselman and Raymond, 1987a, p.177). They concluded that there is considerable scope for improving interview techniques without recourse to complicated theoretical constructs.

Fisher, Geiselman and Raymond (1987a) examined interviews conducted by Police officers in the state of Florida in the United States of America. Eleven tape recorded interviews, conducted by experienced detectives were analyzed. These interviews covered a range of crimes, for example crimes committed with or without a lethal weapon, with one or more suspects, in the day or evening, at the eyewitness' home, in a street or at the eyewitness' place of work. A possible limitation of this study is that so few tapes were used, nevertheless, an analysis of the interviews indicated the following common problems.

1.Interruption of Eyewitnesses' responses.

Fisher et al. found the major problem with the interviews was frequent interruption of the eyewitnesses' responses by the interviewing Police officer. After introducing themselves, all of the interviewers asked the eyewitness to tell them what had happened. However, during this free recall the eyewitness would be interrupted frequently. In the interviews there were, on average, three open-ended questions requiring an extended answer. During the responses to the open-ended questions the interviewer interrupted the eyewitness on average eleven times. In the typical interview the eyewitness was interrupted only 7.5 seconds after they had begun to reply.

They argue that these interruptions cause two main problems. Firstly, they break the concentration of the eyewitnesses when they are trying to retrieve information. The retrieval of information from one's memory is a difficult process at the best of times, but if the eyewitness' concentration is broken by an interviewer's question, then the eyewitness must switch attention from trying to recall information, to the interviewer's question, then back to their memory in order to answer the question. This makes the task much more difficult. Such constant shifting of attention prevents optimal recall of the event. This is particularly unfortunate as free recall typically produces very accurate recall. Further, the increased difficulty of trying to recall information despite constant interruptions may stop the eyewitness from trying so hard to recall information.

The second drawback of interruption is that after eyewitnesses have been interrupted several times they begin to expect to be interrupted throughout the interview. This leads the eyewitnesses to tailor their responses to fit the interview format. As the eyewitnesses expect to have only a short period to respond, they shorten their responses accordingly. Any response which is shortened will not produce as much information, and may exclude information which may be important to the forensic investigation.

Interruptions did not only take the form of questions from the interviewing officers; for example some eyewitnesses were interrupted by the Police officer's radio, and others by someone walking into the interview room - problems which could easily be avoided by turning the radio off or placing a 'do not disturb' sign on the door.

2. Excessive use of question-answer format.

Closely related to the problem of frequent interruption is the excessive use of a question-answer format. Fisher et al. categorised questions as either 'open-ended' questions, where eyewitnesses were required to give a complex response, such as "can you describe the suspects clothing?", or 'short' answer questions that requested a specific answer, such as "What colour was the suspect's shirt?". They found that the majority of questions used in the forensic interviews were of the short answer variety. These questions may have the advantage of eliciting information that the interviewer feels is forensically relevant and prevent the eyewitness from wandering off the point but they can also cause problems.

Fisher et al. found that short answer questioning appeared to produce a less concentrated form of retrieval. Eyewitnesses took less time to respond to short answer questions than for open-ended questions, which may be due (at least in part) to less time being spent actively trying to retrieve information. It was noted that both short-answer and open-ended questions were asked quickly of witnesses, thus there was a short latency between a question answer and the next question, giving no opportunity or encouragement to the eyewitness to elaborate or extend an answer. This use of short answer questions also changes the nature of the task from that of free recall. When short-answer questions are used the interview takes on the format of the interviewer asking short-answer question and the eyewitness giving a brief answer, the interviewer asking another short-answer question, and so on. This means that the interview changes from being directed by the eyewitness to being directed by the interviewer. Fisher et al. comment "It is difficult enough for the eyewitness to retrieve detailed events from memory when actively trying; it is virtually impossible when he remains passive" (p.181).

Using a question-answer format means that all the information elicited is that which is requested. Thus, if the interviewer forgets to ask a certain question, no information in that area is recorded, while if there were unusual occurrences during the crime, of which the interviewer is unaware, questions are not asked and information maybe omitted. As a guide, Fisher et al. suggest that most information should be gathered through the eyewitness's own free recall which should then be followed up with more specific questions later.

3. Inappropriate Sequencing of Questions.

The problems caused by the inappropriate sequencing of questions are similar to those associated with excessive use of question-answer format; both impair recall performance through shifts in attention. Fisher et al. (1987a) noted that many of the questions asked by the interviewers were in a seemingly arbitrary order. They argue that this may impair eyewitness performance through shifting their retrieval efforts from one area to another. For example, if the interviewer asks a visually orientated question about the suspect's face, then follows with an auditory question about the suspect's voice, then returns to a visual target, such as the suspect's clothes, this shift in retrieval attention from one area to another and from one sensory modality to another may impair performance. Indeed, alternating retrieval across modalities has been shown in one study to produce a 19% decrease in eyewitnesses performance (Fisher and Price-Rouch, 1986). Such decrements in performance are not limited to changes in sensory modalities. Shifts within a modality can also cause problems; for example, if one asks a visual question about the suspect's eyes, and then asks about the colour of the ceiling, the eyewitness must shift attention from one visual image to another. A more appropriate technique might be to gather all facial information at one time.

Further problems can be caused by asking what Fisher et al. term 'general knowledge' questions, such as "why do you think he did that?" or " Was he married?", in amongst questions concerning the crime. Shifting from the recall of crime details to general knowledge questions, then back to crime details can cause decreases in the eyewitness's performance.

4. Other problems.

Fisher et al. (1987a) also identified some other problems that did not occur in all of the interviews that they recorded. They considered these problems to be less severe. The problems included negative phrasing, leading questions, inappropriate language, judgemental comments, lack of following potential leads, and an underemphasising auditory cues.

Negative phrasing occurs when questions are asked in the negative form. For example, "you *don't* remember if..?" Phrasing questions in this form may actively discourage the eyewitness from attempting to retrieve information in a concentrated manner. They occurred in many of the recorded interviews. Fisher et al. (1987a) describe leading questions as, questions that subtly suggest that a certain answer is required. Not only are the demand characteristics of the situation likely to produce compliance, but Loftus (1979) has found that leading questions may actually bias eyewitnesses later recollections of an event.

Inappropriate language was found where interviewers used overly formal sentences or words, which were beyond the comprehension of the eyewitness. Such language may not only prevent the eyewitness from understanding the question, but also creates a barrier between the interviewer and the eyewitness which is not conducive to optimal performance. Judgemental comments were occasionally made often about the eyewitness's role in an incident. These may make the eyewitness defensive or may serve to offend the eyewitness, and it is difficult to see how they could enhance recall.

Fisher et al. found that Police officers in their study often failed to follow up on leads that they were given. They cite the case of one eyewitness who described a suspect as looking like a 'newspaperman'. There was no attempt to follow up such comments, to elicit why the eyewitness felt that the suspect looked like a 'newspaperman', that might produce a more objective description. They often found that auditory clues were underemphasised. The Police rarely enquired about what a suspect may have said or if they had an accent.

George (1991) in a field analysis looked at the performance of 28 British Police officers using their usual interviewing procedures. His findings may be considered to broadly support those of Fisher et al. (1987a), thus the issues raised by Fisher et al. would appear to be widely applicable.

The Cognitive Interview

The 'original' Cognitive Interview was an attempt to combine what was known of real life situations and the existing psychological knowledge (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian and Prosk, 1984). The original Cognitive Interview procedure involved four main instructions to the subject or eyewitness.

1. Subjects were asked to *reinstate the context*. They were asked to try to reinstate, in their minds, the context surrounding the target incident. This involved thinking about what the surrounding environment looked like at the scene; for example, rooms, the weather, any nearby people or objects. They were also asked to think about how they were feeling at the time, and their reactions to the target incident.

2. Subjects were asked to *report everything*. They were informed that some people hold back information because they feel that it is not important. However, they were required not to edit their accounts, even if they felt that the information that they remembered was not important.

3. Subjects were asked to *recall the events in different orders*. They were told that it is natural to go over events from beginning to end, but were also asked to try to go through the event in reverse order, or to start with the thing that impressed them the most in the incident, then go from there, working both forwards and backwards.

4. Subjects were asked to *change perspectives*. They were asked to try to recall the incident from the perspective of other people that were involved in the incident. For example, they were asked to try to place themselves in the role of a prominent character in the incident, and think about what they would have seen.

Later, Fisher and his colleagues produced what they considered to be an improved version of the Cognitive Interview (Fisher, Geiselman, Raymond, Jurkevich & Warhafig, 1987b; Fisher, Geiselman and Amador, 1989). This refined or 'enhanced' Cognitive Interview sought to redress certain problems that had been encountered with the original procedure, and to incorporate the findings of the Fisher, Geiselman & Raymond's (1987a) study of Police interview techniques.

An important consideration addressed with the 'enhanced' Cognitive Interview is the structure of the interview. Although the original Cognitive Interview provided instructions at the beginning of the interview little advice was given about conducting the remainder of the interview. Specifically, no guidelines were given about the sequential structuring of the interview. As inappropriate structure and questioning may hinder efficient recall, an important aim of Fisher et al. when modifying the Cognitive Interview was to develop guidelines for the order of forensic interviews.

Many of the refinements to the Cognitive Interview will appear obvious in the light of Fisher et al's (1987a) article described earlier. Essentially the 'enhanced' Cognitive Interview is the standard Cognitive Interview with additional instructions to ensure that the following are included.

1) Time is spent building rapport with the eyewitness. This is achieved by getting to know the eyewitnesses, trying to put them at ease, and ensuring that they are relaxed and aware that they will not be interrupted.

2) The interviewer structures the interview so that it is directed by the eyewitness, thus allowing the eyewitness time to concentrate, and structuring the interview so that it is "compatible with the mental operations of the witness" (Memon & Bull, 1991, p.295). The interviewer avoids fixed styles of questioning, tries to empathise with the eyewitness' mental operations and avoids interrupting the eyewitness by holding back questions where appropriate.

3) The interviewer helps the eyewitness to produce 'focussed retrieval'. The interviewer must "encourage and assist the witness to generate focussed concentration" (Fisher, Geiselman & Amador, 1989, p.723), principally through motivation. Further, Fisher et al. (1989) recognised that retrieval is a difficult task that requires motivation; thus they explicitly state that "the effective interviewer must encourage the witness to make the extra effort" (p.723).

Thus although the 'enhanced' Cognitive Interview is similar to the original Cognitive Interview, it differs mainly in that the eyewitness directs the content and direction of the interview rather than the interviewer. In this respect it also differs from conventional interviews.

Empirical support for the Cognitive Interview

Some initial support for the Cognitive Interview came from a study by Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian (1984). This was an experimental investigation of the effectiveness of the original Cognitive Interview compared to control conditions. The study used 16 undergraduate psychology students, who were told that they would be taking part in an experiment to improve their memory. However, while apparently taking part in a memory experiment they witnessed a staged incident; this was the interruption of the 'experiment' by an argument. Subjects were tested for their recall of the argument, using a test booklet. The initial question was "..what do you remember of the incident involving the person (or people) who interrupted the experiment at our last meeting?"(p.76) Subjects were asked to put one piece of information on each line of the test booklet. They were asked to rate their confidence in each piece of information that they provided. Subsequently more "pointed" (i.e. specific) questions were asked.

The experiment revealed that significantly more correct information was recalled with both the open, and the "pointed" questions, with the Cognitive Interview. Moreover, this increase in correct information was achieved without a corresponding increase in incorrect information. Eyewitness confidence in correct information was significantly higher in the Cognitive Interview than in the control interview. Eyewitness confidence in incorrect information was also higher in the Cognitive Interview condition, but this was not significantly greater than that for the control conditions; it can be noted, however, that only a limited number of subjects were used, and there were few incorrect responses.

Aschermann, Mantwill & Kohnken (1991) conducted a partial replication of the Geiselman et al. (1984) study. This study used 29 German undergraduate psychology students. Subjects were shown a short film which was presented incidentally. They were tested between two and nine days later. Subjects were tested in a similar manner to the Geiselman et al. (1984) study. The results showed a significant overall increase in correct information recalled in the Cognitive Interview condition, compared to a standard interview condition, and this was especially apparent with the open-ended initial question. There was no significant difference between the Cognitive Interview and standard interview in the amount of incorrect information produced although there was a trend for more incorrect information to be produced with the Cognitive Interview, especially with open-ended questions. Unfortunately, Aschermann et al. did not take any measures of confidence so it was not possible to estimate whether confidence in incorrect answers is increased by the Cognitive Interview.

It may be of some importance that both of the above experiments differed from a standard Police interview, in that rather than the interviewer listening to the eyewitness, then writing a statement, the subjects effectively wrote their own statements. This may have been advantageoud in that the subjects may have felt less hurried and more able to think about and report details.

Geiselman, Fisher, MacKinnon & Holland (1985) conducted a follow-up experiment to their original study, using a larger sample of 89 undergraduate psychology students. Subjects were presented with one of four films that were used by the Los Angeles Police Department to train Police officers. In each film at least one individual is shot and killed. Subjects were made explicitly aware that they would be later be tested on the film.
After approximately forty-eight hours subjects recall for the film was tested. In this study interviews were conducted by law-enforcement personnel trained in the use of the standard Cognitive Interview, 'hypnosis' or using their own usual interviewing methods. By comparing the Cognitive Interview with the law-enforcement personnel's usual methods this study avoided the comparison of the Cognitive Interview with an artificial 'control' procedure devoid of many of the facets identified by Fisher et al. (1987).

By transcribing the information produced by the subjects, it was found that there was a significant 40% improvement in the amount of information produced using the Cognitive Interview when compared with the standard interview, and a 30% increase with the use of hypnosis compared to the standard interview. There was no significant difference between the amount of incorrect information produced by the groups. This is noteworthy given that there is considerable concern over the usefulness of 'hypnotically' elicited information because of problems relating to increased confidence in inaccurate information, increased incorrect information, and confabulation (Wagstaff, 1989; Smith, 1983).

Other experimental investigations have shown a similar, though variable, pattern of increases in recall for the original Cognitive Interview with no significant increases in errors (Bekerian & Dennett, 1993; Geiselman, Fisher, MacKinnon & Holland, 1986; Kohnken et al., 1991; Memon and Bull, 1991).

More recently, George (1991) conducted an experiment using the 'enhanced' Cognitive Interview. Subjects in a lecture witnessed a staged incident. Two weeks later subjects were interviewed by Police officers using a) the 'enhanced' Cognitive Interview, b) a procedure of Conversation Management (a technique designed "to equip interviewers in the social and communication skills required to open, and keep open, channels of communication in order to find out facts" p.3.), or c) standard Police interviews. The results of these interviews were transcribed- interviewers were explicitly told " There is no need to capture the information in writing" (p. 97). There was a trend for the 'enhanced' Cognitive Interview to produce more information, but this was not statistically significant (this could have been due to the small sample size; only four subjects in the Cognitive Interview group and 15 in the whole experiment). There was no indication increased errors or confabulations in the Cognitive Interview group.

In addition to studies comparing the Cognitive Interview with various controls, Fisher, Geiselman, Raymond, Jurkevich & Warhaftig (1987) also conducted an experimental investigation comparing the Cognitive Interview with the 'enhanced' Cognitive Interview. A similar protocol to that of Geiselman et al.'s original 1984 study was used. Subjects were shown a video recording of a simulated violent crime then their memory for the video was tested 48 hours later. Subjects recall was transcribed from tape recordings of the interviews. The study showed the 'enhanced' Cognitive Interview to produce 45% more correct information than the standard Cognitive Interview, an increase which was not accompanied by a corresponding increase in errors. Fisher et al. further coded the data produced by the 'enhanced' Cognitive Interview and the standard Cognitive interview, to determine if the extra information produced by the 'enhanced' Cognitive Interview was simply trivial information. They found the increased information produced by the 'enhanced' Cognitive Interview' to have a similar proportion of information relevant to the crime as the standard Cognitive Interview. However, although this extra information recalled in the 'enhanced' Cognitive Interview was deemed **relevant** to the crime, on the basis of the coding scheme as reported, it is not clear whether this information would was more, less, or of equal **use** to the forensic investigation than information produced using a standard Cognitive Interview procedure.

The Cognitive Interview with children

Experimental investigations of the Cognitive Interview have also included various evaluations of its efficacy with children. The use of child witnesses has become a controversial issue recently because although children may potentially be able to provide useful information, though their suggestibility and accuracy have been questioned. Clearly, therefore a technique to improve their memory performance in an accurate would be beneficial.

Geiselman & Padilla (1988) used the original Cognitive Interview to test the memories of children, between seven and twelve years of age, for a video they were shown of a liquor store hold up. Their ability to remember details was tested three days later using either the Cognitive Interview or a standard interview. Children interviewed using the Cognitive Interview produced 21% more information than those tested with the standard interview without a corresponding increase in inaccurate information.

Saywitz, Geiselman & Bornstein (1992) conducted a similar experiment but using a staged incident rather than a video film. They found similar effects; a 20% increase in information recalled for eight and nine year olds using the Cognitive Interview compared a control group of similar age. A 44% improvement was found for 11 and 12 year olds, when compared to the appropriate control group. This performance was further improved to increases 25% and 66% respectively, over controls in the corresponding age groups, if the Cognitive Interview was practised before the test session. No increase in the amount of incorrect or confabulated information was observed.

Dietze & Thomson (1993) compared the recall performance of six year olds, 11 year olds and adults, both with and without an abbreviated form of the Cognitive Interview. Their results showed an increase in the amount of information recalled with the Cognitive Interview when compared with the free recall condition for each age group. No significant corresponding increases in errors were found. The amount of information recalled also increased with age. Interestingly, Dietze & Thomson (1993) suggest that children's failure to perform as well as adults, is due to them encoding less information, so even with an optimal retrieval strategy they would not perform as well as adults. Nevertheless they suggest:

> "If the performance differences between children and adults only reflect children's problems in utilizing an appropriate retrieval plan, then one would expect that children would benefit more from the use of an appropriate retrieval plan than would adults" (p.105).

They go on to state that evidence suggests that the Cognitive Interview is an appropriate method of providing retrieval cues for children as well as adults.

Problems of ecological validity

An obvious critism, however, of many the Cognitive Interview studies is that they lack ecological validity; i.e. they are too artificial. Geiselman et al. (1985) make claims about their stimulus materials such as, "The scenarios are realistic in that monitored reactions of officers in training have been found to be comparable to reactions that would be expected in similar street situations" (p.404). A thorough evaluation of this assertion would require further information about the empirical work that forms the basis of these conclusions. It is not obvious, for example, that a film clip showing a killing will produce similar reactions to a real-life killing. Also, unlike in some of experimental studies, in 'real' life situations, of course, eyewitnesses are not usually aware that they may be tested later, so they would be unlikely to encode the information regarding the target event as effectively as in a situation where testing is anticipated.

Fisher et al. (1989) have noted that the results of their experiments would always be questioned until they are demonstrated in the real world. They state "...if the Cognitive Interview is to be applied outside the friendly confines of the laboratory, it must be demonstrated to be effective in the real world" (p. 724). Fisher et al. (1989) therefore set out to investigate the 'enhanced' Cognitive Interview in a field setting. Also, shortly afterwards, George (1991) conducted a similar field investigation.

The Cognitive Interview in the field

On first consideration, the Fisher et al. (1989) and George (1991) studies appear to provide some of the most convincing evidence to date of the effectiveness of the Cognitive Interview.

The method used in the field study by Fisher et al. (1989) was to tape record interviews of eyewitnesses to real crimes. The interviews were conducted by 16 experienced detectives from the robbery division of a Police force in Florida, USA. Preliminary recordings were made of interviews conducted by the detectives on eyewitnesses, before any training in the Cognitive Interview was given. In all, 88 interviews were recorded before training, and 47 interviews were conducted after training; 24 in a group using the Cognitive Interview and 23 in a control group.

The interviews that were used were selected according to the following strict criteria. 1) The case had to be severe enough that time would be made available to conduct a thorough interview. 2) The eyewitness must have had a 'decent' chance to observe the incident and the suspect. 3) The eyewitness had to be fluent in English and also co-operative.

Fisher et al. (1989) state that some interviews were rejected as unsuitable because of reasons such as, the eyewitness was intoxicated, the interview was a couple of days after the incident, the suspect was known to the eyewitness, or a suspect had been detained for identification. The last reason may appear strange but apparently when a suspect is in custody, Police tend to take a less comprehensive attitude to interviews, preferring to secure a positive identification instead. (Although rejection for these reasons would clearly create a more homogenous sample of interviews; in the case of intoxication or delayed recall one might have expected the Cognitive Interview, with its emphasis on reinstating different context, to perform even better than in the other conditions.) Seven detectives completed the training programme. The recording of post-training interviews took seven months to complete and the tape recordings were then transcribed by research assistants.

Fisher et al. (1989) evaluated the effectiveness of the Cognitive Interview in two ways; 1) by comparing the number of 'facts' elicited before and after training in the use of the Cognitive Interview, and 2) by comparing the number of facts elicited by the trained detectives using the Cognitive Interview and the control group of detectives who were still using standard techniques. When detectives who were not going to be trained in the Cognitive Interview were compared to the detectives who were to be trained in the Cognitive Interview there were no significant differences between the two groups. However, after training in the Cognitive Interview there was a significant improvement; 63% more information was recalled by eyewitnesses interviewed by detectives trained in use of the Cognitive Interview compared to those interviewed by the 'control' detectives. Moreover, detectives in the Cognitive Interview trained group showed a 47% increase in the amount of information that they elicited from eyewitnesses compared to their previous performance before training.

However, whilst these results seem impressive, there are a number of difficulties in the interpretation of this study. One possible problem concerns the way in which statements were scored. Fisher et al. did not score opinionated responses, such as comments like "the guy seemed nervous" (p.724). It is possible that that this scoring method may have lead to artificially high performance in the Cognitive Interview group. Because of the explicit instructions to try harder (and perhaps implicit instructions to be more confident), comments which would not have been scored such as "the guy seemed nervous" in the Police interview may have changed to a comment like "the guy was nervous" with the Cognitive Interview and have been scored accordingly.

It is also notable that of the seven detectives trained in the Cognitive Interview, one (10%) produced a **decrease** in performance of 23%. Clearly, on the basis of such a small sample it is difficult estimate whether this was a curious anomaly or whether this represents a potential problem. Fisher et al.

(1989) comment of this detective: "Not coincidentally an analysis of the post-training interviews showed that he was the only one of the seven detectives who did not incorporate the recommended procedures into his post-training interviews,"(p.724). However, despite this explanation, this example does raise further questions. How and why did this Police officer fail to incorporate the recommended procedures into his interviews? Presumably, he passed the training procedure, including a practise interview in the field, and received individual feedback on his performance, as outlined in Fisher et al's (1989) method section. Did his performance improve or become worse over the repeated attempts? As such a result has not been reported in previous studies in the laboratory or experimental situations why did it occur in the field? This example suggests that if the Cognitive Interview is to be used widely, the performance of individual interviewers should be carefully monitored. Indeed, George (1991) has also noted that some officers pick up the Cognitive Interview better than others, and suggests that training should be concentrated on certain individuals.

Another problem of interpretation concerns how Fisher et al. (1989) tackled the question of accuracy. Obviously, in a field situation accuracy is difficult to determine as there is often no way of definitively establishing what actually occurred. Fisher et al. (1989) therefore estimated accuracy by comparing each eyewitness report with that of what they term another 'reliable' source, when this was possible. In 22 cases this source was another eyewitness, in one case a confession, and in one case information was supplied by a video camera. Fisher et al. (1989) found there to be a 93% corroboration rate with information produced by a 'reliable' eyewitness for information produced by detectives untrained in the Cognitive Interview, and a 94.5% corroboration rate for detectives using the Cognitive Interview. Fisher et al. (1989) note that their corroboration levels ares high when compared with the accuracy levels typically produced by laboratory studies, and they cite the similar findings of Yuile & Kim (1987). They state:

> "If this difference between laboratory and field studies continues to appear, one may question the validity of describing in court the accuracy rates found in the laboratory as evidence of the general unreliability of eyewitness testimony in field cases" (p.725).

However, corroboration rates in field studies and estimates of memory accuracy in experimental studies may be quite different measures. There could be a marked difference between the accuracy of corroborated information and the accuracy of all information produced. By definition corroborated information is information which two or more eyewitness have recalled. Therefore one can assume such information is probably *central* to the eyewitness situation and more likely to be recalled by several eyewitnesses. Laboratory situations typically consider both central and less important *peripheral* information to determine accuracy rate. If only the *central* information from laboratory studies were considered, the accuracy rate would probably be closer to that of the corroboration of the Fisher et al. (1989) and Yuile & Kim, (1987). Also as Fisher et al. (1989) themselves note, just because two items are correlated by two eyewitnesses, does not necessarily mean that they are accurate -both may be wrong.

However, perhaps the most important problem concerns how we identify the elements in the Cognitive Interview responsible for the reported improvements. For example, as already pointed out, the group trained in the Cognitive Interview produced a significant increase of 47% more information using the Cognitive Interview, compared to their previous performance. Of the seven detectives, the range of improvement, for six of them, was between 34-115%. What exactly was responsible for this range? How did the detective who produced a 115% increase in the number of facts elicited manage this? Did he reinstate context better than the others? Was his pretraining performance especially bad due to excessive interruptions? Did the training motivate him to try harder to motivate his own eyewitness? Without answers to such questions it is difficult to determine whether the improvements were a consequence of the Cognitive Interview per se, or to more generally features of the situation such as 'training' per se, or the fact that the situation was novel or different. The detectives were aware that they were evaluating a new technique, which they had been specially trained to use. The training for, and use of, a new technique may in itself have produced positive improvements in the Police officers performance, by motivating them to try harder and in turn, to motivate the eyewitnesses to try too. Whilst it could be argued that this effect may be desirable no matter how it comes about, it may fade as the 'novelty' of the technique wears off, and may have implications for what training should entail.

Of relevance here therefore is the field investigation in the United Kingdom by George (1991). In this study, 28 Police officers were evaluated in one of four conditions; seven in each condition. A recording of an interview performed by each officer was evaluated before each was trained in an interview technique or placed in the control group. The interview techniques were: 1. the 'enhanced' Cognitive Interview; 2. Conversation Management; 3. Conversation Management combined with the 'enhanced' Cognitive Interview and 4., a control group. The results indicated that the 'enhanced' Cognitive Interview showed an improvement when compared to the standard Police interview control group of 14% more information. When compared to performance before 'enhanced' Cognitive Interview training this improvement was 55%.

This advantage was for all kinds of information (i.e. who, what, when, where, how and why). Neither the Conversation Management nor the combination of Conversation Management and 'enhanced' Cognitive Interview produced more information than the untrained group. These results would suggest that it was not 'training' per se, or novelty alone that accounted for the improvements that occurred with the cognitive interview.

Nevertheless, a major finding in George's study was, that of the four mnemonic strategies suggested in the cognitive interview, three were hardly utilised. The instruction "not to edit anything out" was only minimally employed. The two other mnemonic aids which were rarely used were the instructions to "change of perspective" and a "change of order". George notes that it is unsurprising that officers rarely used the change of perspective mnemonic as, "it is not an easy concept to ask someone to put themselves in someone else's shoes to review an event asking them to say what they think they would have seen, and remain confident that there will be no confabulation" (p.117). Critics have also suggested that the use of the "change of perspective" mnemonic may make it difficult to use such statements in court, especially if children are interviewed, because of a danger of confabulation. The Police officers in this study may have had an intuitive grasp of this and so, frequently did not use the technique. Some research into why these three techniques were not used and the implications for their inclusion in the Cognitive Interview would seem is appropriate. It is not clear if a similar pattern of mnemonic usage was present in the Fisher et al. (1989) study.

The fourth mnemonic, reinstatement of context, was widely used, apparently to great effect. This lead George to conclude that "where contextual reinstatement was present more information was elicited regardless of question type" (p.118). This effect was consistent across the seven types of information into which George coded recall. Reinstatement of context thus appears to be the most reliable technique for increasing recall. Why, however, the effects of the Cognitive Interview should be eliminated when it is combined with another procedure (Conversation Management) remains somewhat of a mystery.

Although it is time-consuming to perform, the field study clearly has been an effective paradigm in this area. And despite the problems in interpretation, it is very notable that the improvements found in field studies are remarkably similar to those which have been found in previous laboratory based experiments. Indeed, it could be argued that the Cognitive Interview might work *more* effectively in the field than in artificial experimental situations, because of greater contextual disparity between encoding and recall contexts in real life. In the laboratory, both encoding and retrieval contexts are usually fairly similar;

for example, studies are conducted in college rooms in a situation inducing low emotional arousal. In real eyewitness situations the differences between encoding and retrieval conditions are likely to be markedly different. For example, an individual may be drunk and frightened in a street at the encoding of an event but sober and relaxed in a Police interview room when they are required to remember the event. Assuming the mental context reinstatement is effective, one might therefore expect the Cognitive Interview to show even greater performance increments over standard interviews in these latter circumstances.

Application of the Cognitive Interview to U.K. Police procedures

Nevertheless, even if we assume that the Cognitive Interview is a relatively useful technique for memory enhancement, how feasible would it be for the Police in the United Kingdom to employ such a procedure?

The methods used in the Fisher et al. (1989) and George (1991) field studies are in some ways considerably different to those used by United Kingdom Police. Usually the Police do not have transcribers to record everything that an eyewitness says. Typically in the United Kingdom, the interviewer takes notes as the eyewitness recalls information. At the end of the interview the interviewer writes a report in the first person, as if the eyewitness had written the report themselves. The Police officer then reads the statement to the eyewitness and asks him or her to sign it. This procedure is likely to produce errors for a number of reasons that may limit the effectiveness of a Cognitive Interview.

1. The interviewer may be unable to remember all that the eyewitness has said so may have to 'guess' some of the information.

2. Interviewers may select or alter information from the account, to make the statement fit with their own preconceptions of what occurred, and their own stereotypes of people's behaviour, whilst omitting information that contradicts this.

3. Interviewers will produce the statement in their own words. These words may be either more simple or more complex than the eyewitness. When unable to spell words the Police officer may be tempted to use simpler words with subtly different meanings. Thus the eyewitness's statement may be subtly changed.

4. Any errors produced in a statement are likely to influence an eyewitness' subsequent court performance. As court appearances are often a considerable time after a crime occurred eyewitnesses are often read 'their' statement in order to refresh their memory for the incident. If eyewitnesses' memory of an incident has become vague they are unlikely to question the accuracy of 'their' statements.

5. Although eyewitnesses are asked if they are happy with their statements before signing them, in the social situation where a Police officer has spent a considerable time writing a statement, it may be very difficult for an eyewitness to ask them to change it especially if they feel that their statement requires changing dramatically, and so a great deal of effort by the Police officer.

The way in which information from eyewitnesses is recorded by the Police, for both operational and court use, is clearly of considerable importance yet there has been little research conducted in this area. If the standard procedures are maintained alongside the Cognitive Interview then in real-life situations there may be additional difficulties. The use of the Cognitive Interview will mean that there will be long periods without interruptions. During these periods the Police officer will likely to be unable to take notes quickly enough to record all that an eyewitness says, and will therefore rely more on his or her memory of what the eyewitness said. Here we may find a 'Catch 22' situation, the standard Police interview does not produce as much recall as the Cognitive Interview, but what is produced can be noted by the interviewing officer, whereas the Cognitive Interview produces greater recall but the Police officer is unable to remember all the information.

The area of Police officers note-taking must therefore be effectively evaluated. As has been outlined, it is of considerable importance that eyewitness testimony is accurate and comprehensive but also that police officers can record and use such information. Fisher & Geiselman (1992) suggest that interviewers "need to develop some type of shorthand method of note-taking" (p.82), as eyewitnesses talk faster than interviewers can write. They go on to suggest that: "Investigators within a department simply pool their personal abbreviations into a master list that everyone can share" (p.82).
However, if the use of the Cognitive Interview in practical situations is such that a whole shorthand procedure must be learned then this will add considerably to the training time necessary to use the technique.

Possible methods of circumventing these difficulties might include asking the eyewitness to speak slowly or to asking eyewitnesses to write in their own words what occurred. In the United Kingdom illiteracy is comparatively rare, so requesting that subjects write their own statements is feasible, although eyewitnesses would have to be highly motivated by the investigating officer so that they do not give up too easily, as many might feel that writing is a tedious task. A more time consuming alternative might be to transcribe or take detailed notes of tape-recorded interviews.

Another important practical issue concerns training. According to the experimental reports it appears that the 'original' Cognitive Interview technique can be rapidly picked up by prospective interviewers. Unfortunately Fisher et al. (1987b) do not specify how long the training of interviewers took in the initial test of the 'enhanced' Cognitive Interview, though they mention that their interviewers had no previous training. In their field study, Fisher et al. (1989) report that, although one officer apparently did not grasp the concepts correctly, only approximately five hours was required. Fisher (1993) says, "There really is no trick about doing this type of interview. You just have to sit there and listen." Though he adds, "it takes quite a lot of training" to get interviewers to do this (p.28). Officers in the George (1991) study took part in a two day training course before using the 'enhanced' Cognitive Interview. This short training time might become longer, however, if alternative methods are required to record information.

Nevertheless, no matter how effective the training or technique, it may still be the case that the Cognitive Interview is inappropriate for certain kinds of investigation. In fact, the problems with conventional interviewing techniques may not actually be as clear cut as as they at first seem. For instance, informal discussions with Police officers suggest that there are specific reasons for officers to use interruptions and a question/answer format. Some Police officers say that they will use such strategies deliberately to curtail an eyewitness's report in certain situations (i.e. to limit the eyewitness's report only to that which the officer deems necessary). One such situation, might be when an officer has to interview a number of eyewitnesses. If the officer is at a crime scene and is confronted with a number of witnesses who require interviewing, he or she must interview each quickly to ensure that all are interviewed. If a long time is spent interviewing each witness the officer risks other potentially valuable witnesses leaving the crime scene. Thus, a method of conducting a rapid, succinct, interview can sometimes be more useful than a technique which produces a lengthy, more complete account. A similar situation, where a brief interview is necessary is when a Police officer has already has been given another crime to deal with after a current one and so has to rapidly respond to that request. Thus, the above factors evident in many Police interview procedures may not always be as disadvantageous as Fisher et al. (1987a) have suggested; it may depend on the situation.

When considering Police interview techniques it should also be noted that Police officers do not just interview victims and bystanders, they also spend a considerable time interviewing suspects. This may impact on the interviewing techniques which they use. For example, asking a closed question which requires a rapid answer may not produce the best testimony from an eyewitness because it does not encourage 'focused' retrieval and an elaborated response; but such a technique may lead a suspected criminal to make an incriminating reply precisely because he or she was not given the time to think of a false statement to give. Thus the requirements of a 'good' interview of a suspect may be considerably different to those of a 'good' interview of a non-hostile eyewitness. Some Police officers report that if they alternate from interviewing a suspect to interviewing an eyewitness in a short period of time it is difficult to switch quickly from the frame of mind necessary to conduct a 'good' interview of a suspect to that necessary to conduct a 'good' interview of an eyewitness. Thus, they find themselves interviewing eyewitnesses in an inappropriate manner. This may explain some of the problems identified by Fisher et al. Further it may have practical implications for Police interviewing procedures; perhaps some Police officers should specialise in eyewitness interviewing while others specialise in the interviewing of suspects.

George (1991) noted that Police officers who had not been formally trained in interviewing techniques "mysteriously share a common schema for deriving information". He states:

> "In the absence of training... Police officers somehow all perform and acquire information in the same manner. Intuitively, as experienced practitioners, it was tempting previously to assume that police necessarily perform this task in the most effective manner possible" (p.125).

These standard techniques are similar to those outlined by Fisher el al. (1987a). The fact that a 'common schema' exists may itself indicate something of importance. Perhaps, on the basis of experience, this schema has proved on balance to be very effective; that is to say, the fact that many Police officers use a similar style may indicate that standard Police techniques may have considerable utility. Given the practical considerations just identified, standard Police interviewing techniques may represent a good *compromise* interview method; i.e. a useful general framework with which to 1. interview suspects, 2. interview eyewitnesses (sometimes in a limited period of time) and 3. be able to take appropriate notes. Moreover, it may suit resourcing policies; there seems little point in gathering a large amount of information for a minor crime if there are not enough resources to follow up leads anyway. However, for certain *specific* situations it may not be the most appropriate method. For example, when a serious crime has been committed more resources, especially time, are usually allocated. In such instances it maybe more appropriate to use a technique which may take longer but generate more information, such as the Cognitive Interview.

Another important practical point concerns the appearance of eyewitnesses who have been interviewed using the Cognitive Interview in court. The appearance of eyewitnesses who have been 'hypnotised' in the hope of improving recall is a topic of considerable and vigourous debate (Wagstaff, 1989), such that the Home Office has issued strict guidelines on the conduct of interviews using hypnosis (e.g. the whole thing must be videotaped). It may be the case that the Home Office will advise similar guidelines for the conduct of Cognitive Interviews.

Conclusions

At the beginning of this chapter three criteria were outlined by which the Cognitive Interview might be evaluated. To what extent does the Cognitive Interview satisfy these criteria?

1. Does the technique increase reliable recall, and 2) does the technique produce testimony in which confidence is related to accuracy?

Although there are also some possible problems with the methods of scoring responses which could have inflated Cognitive Interview performance, virtually all of the studies described here have shown a substantial increase in recall with the Cognitive Interview compared to control conditions. Also confidence in correct answers has been significantly increased. No significant increases in incorrect information, or confidence in incorrect information, have been reported, though there have been some non-significant trends in this direction.

It can be noted here that when 'hypnosis' was compared with the Cognitive Interview there was a significant increase in the amount of information recalled without corresponding increases in incorrect information for both 'hypnosis' and the Cognitive Interview (Geiselman et al. 1985). However, hypnotic memory enhancement procedures have been plagued, not so much by the problem of producing incorrect information per se, as the problem of inducing false confidence in inaccurate responses in response to leading questions (e.g. Smith, 1983; Wagstaff, 1989). The effects of providing leading questions using Cognitive Interview procedures have yet to be thoroughly investigated. Fisher et al.'s interviewers have tended to ask very few leading questions so it is difficult to assess what would have if many such questions had been asked. The interviewer under the watchful eye of the psychologist may be rather more cautious in asking leading questions than the investigator in everyday life desperate for leads or a conviction. Thus until more extensive research is conducted that looks specifically at the 'false confidence' issue in relation to leading questions, it is not necessarily possible to rule out similar problems with the Cognitive Interview.

However, on the whole, until evidence to the contrary is presented, it seems reasonable to conclude that the Cognitive Interview has the potential to increase accurate recall.

3. Is the technique easy and practical to implement?

The empirical evidence suggests that the technique can be taught in a relatively short period of time, approximately four hours for real forensic situations. However, researchers seem to have have paid little attention to how Police officers take down and use information. Officers may need further training in this area if the Cognitive Interview is to operate successfully. Also, perhaps more important, practitioners should not be too zealous in their dismissal of conventional interview techniques; in some cases the Cognitive Interview might be overly complex and blatantly counterproductive as an interview method.

But perhaps most perplexing for researchers is the issue of exactly what it is about the Cognitive Interview that is responsible for the increases in recall. How much of the effectiveness of the Cognitive Interview is due to its 'cognitive' components and how much is simply due to it increasing the motivation of eyewitnesses and interviewers to perform well? Are all, indeed any, of the mnemonic strategies in the Cognitive Interview necessary? Significantly, more recently, Fisher (1993) has commented:

> "My impression is that [the Cognitive Interview] works primarily because it facilitates communication and only secondarily because it improves memory retrieval" (p.28)".

Future research might therefore usefully address the importance of eyewitness' perceptions concerning how well the Police officer interviewed them.

Indeed, the interview of an eyewitness provides a good opportunity to enhance Police-public relations (Fisher & Geiselman, 1992). A well conducted interview with a concerned and genuinely interested Police officer is likely to increase the eyewitness's respect for and future willingness to co-operate with the Police. The alternative, a poorly conducted interview with an unconcerned Police officer, showing little interest in what the eyewitness says, may discourage the eyewitness from future support of the Police. As a crime incident is likely to be a talking point for a long period of time, then the eyewitness's impression of the Police is liable to be relayed to a large number of people. Thus it may have a large impact on a locality. Perhaps here then is an unanticipated benefit of the Cognitive Interview that may prove to be its most positive contribution to forensic investigation.