SOME INFLUENCES OF COURTROOM QUESTIONING STYLES ON EYEWITNESS CONFIDENCE AND ACCURACY

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by Jacqueline Mary Wheatcroft.

March 2002

ACKNOWLEDGEMENTS

Firstly, I would like to take this opportunity to thank my supervisor Graham Wagstaff for his excellent and consistent supervision and support for the duration of my scholarship. Secondly, I would like to express thanks to other people without which this thesis would probably never have been completed. These include: Mark Kebbell, Par Anders Granhag, Shane Johnson, Peter Banister, and the Secretarial and Technical staff of the Psychology Department at Liverpool. In addition, my mother, my husband Mitchell, his mother and father, and my children James, Daniel, and Charlotte.

This PhD is dedicated to my father John Boles who died suddenly during the writing stage on 29th June 2001.

I would also like to thank the University of Liverpool and the Department of Psychology for funding this PhD scholarship.

ABSTRACT

Once in court, witnesses are subject to a number of procedures that aim to elicit testimony, but none is so important as cross-examination. For the witness, cross-examination represents an inquisition that allows the advocate to employ a range of tactical procedures such as linguistic variation, negative feedback, and the targeting of critical items, that are articulated through the use of leading and pre-suppositional questions (i.e. collectively known as a 'lawyerese' questioning style). Yet, despite the large psychological literature on eyewitness accuracy and confidence, little empirical research has been conducted on performance under cross-examination conditions, and individual differences in the abilities to give accurate and credible testimony to the court. In view of these considerations, the central focus of this thesis was to investigate how ordinary witnesses perform under cross-examination conditions. Four main questions were investigated. 1) Do cross-examination questioning styles, including the provision of negative feedback, affect witness confidence and accuracy, and the relationship between confidence and accuracy (C-A)? 2) Do accuracy, confidence, and the C-A relationship, so influenced, remain stable over time? 3) Do jurors assess the accuracy and confidence of witnesses differentially according to the way in which they are examined? And, 4) Can we predict who will be a good and poor witness under cross-examination? Results showed that a lawyerese style alone had little effect on accuracy generally, but a negative effect on confidence accuracy calibration for difficult items, whereas addition of subtle feedback improved confidence accuracy calibration, but at the expense of accuracy for difficult items. A delay of 6 months between stimulus presentation and questioning reduced accuracy generally, but, in some cases, confidence accuracy relationships actually improved. Results also showed that juror/observers tended to link witness confidence with accuracy, however, a lawyerese questioning style with subtle negative feedback style destroyed this link. Furthermore, presenting the testimony of a poor witness first increased observer's perceptions of that witness' accuracy, confidence and competence. Attempts to find a predictor of performance under examination were largely unsuccessful with the possible exception of verbal fluency, a neuropsychological test of brain function. It is concluded that, if these results are generalisable, many of the assumptions made by some legal professionals about the efficacy of cross-examination procedures should be seriously questioned.

PREFACE

The legal system is a structured, institutionalised forum for handling criminal or civil conflict. However, it is important to be aware that decision-making does not occur solely as a result of the mechanical application of rule or court procedure, but the eyewitness, particularly in criminal cases, represents a crucial component of the decision process. All witnesses, whether victims or other persons at the scene of the crime, function as data-gathering sources whose construction of reality can significantly affect the fate of the defendant, thus having the potential to impact greatly on the Criminal Justice System.

There are many varieties of evidence which may be presented to jurors and jurists in a criminal court. However, the principle of orality is a traditionally claimed virtue of the common law trial, based upon the belief that the credibility of evidence can only be assessed if the court observes the witness utter it from the witness box. Clearly then, those who are 'witnesses to the facts of the case' play a centrally important role in most investigative and legal processes. Once in court, however, the witness will be subject to a number of procedures aimed at eliciting testimony, but none so important as that designed to test witnesses' veracity: cross-examination.

All examinations to which the witness is subject to in court may be classed as interviews. However, some interviewing procedures, both in and outside of the courtroom, are necessarily interrogative. For example, the 'standard police interview' requires that the process uncovers as many relevant facts relating to the investigation as possible; while 'cross-examination' aims to elicit further facts favourable to the cross-examiner. Cross-examination

remains unique then in that such a procedure is an interrogation both in context and nature. That is to say, it consists of a range of questions put closely and formally aimed specifically at directly challenging the substance of witness evidence given, and to attack the credibility of the witness. Thus, the cross-examination is an inquisition focused on the witnesses and what they know, rather than being merely inquisitorial about what they might have observed, as in the police interviewing ordinary witnesses to an event. Furthermore, because cross-examination is designed to probe the accuracy of evidence-in-chief, and to expose dishonest or unreliable evidence, leading questions are always permitted, and a range of other tactics, techniques and styles may be employed. Such techniques and tactics include; linguistic variation, negative feedback, and critical item difficulty. To date, however, no empirical research has been published on the effects on testimony of interview styles, or negative feedback, as might be used by examiners in a courtroom.

This is important given that some have expressed concern that court officials lend too great a weight to evidence given by witnesses when reaching decisions, believing them to be accurate and thus truthful. Accuracy appears to be generally inferred from the overall credibility of the witness, which is determined by a range of factors: the inherent consistency of evidence (e.g. whether the witness appears to be suggestible); corroboration, where possible; performance in witness box (e.g. whether there is a witness defect rendering evidence unreliable); observation of physical manifestation of truthfulness or mendacity, and the inherent probability or improbability of evidence. One of the most influential measures used by courtroom officials to determine accuracy is the confidence expressed by witnesses in respect of their

testimony. However, as yet we know nothing about how witness confidence and accuracy, and the relationship between confidence and accuracy (C-A), may be mediated by the styles, tactics, and procedures engaged in during cross-examination.

Although much academic and legal debate has centred upon cross-examination, most arguments have been put forward by linguists or those interested in the welfare of those who might be susceptible giving their oral evidence in court (i.e. children, those with learning difficulties, and some older adults). This focus has tended to separate the 'ordinary' witness from those witnesses who are classed as vulnerable. The focus of research upon these groups of witnesses has lead to a virtual vacuum of empirical investigation into cross-examination and its effects upon ordinary witnesses; i.e. those who are statistically much more likely to be called upon to give evidence in court. Furthermore, it is likely that potential 'ordinary' witnesses are not a homogenous group, and they may vary in terms of their abilities to give accurate and credible testimony to the court under interrogative conditions.

These considerations give rise to a number of questions that are addressed in the present thesis. 1) Do cross-examination questioning styles, including the provision of negative feedback, affect witness confidence and accuracy, and the relationship between confidence and accuracy (C-A)? 2) Do accuracy, confidence, and the C-A relationship, so influenced, remain stable over time? 3) Do jurors assess the accuracy and confidence of witnesses differentially according to the way in which they are examined? And, 4) Can we predict who will be a good and poor witness under cross-examination?

CONTENTS

TABLES

FIGURES

APPENDICES

PART 1

LITERATURE REVIEW AND INTRODUCTION TO THE EXPERIMENTAL WORK

CHAPTER 1

EYEWITNESS TESTIMONY AND THE COURTROOM

1.1 The importance of eyewitness testimony

General interest in the field of eyewitness testimony has been considerable since Whipple's first reports and enquiries into the subject dating back to the beginning of the century (1909, 1910, 1917). However, most of this early work failed to recognise fully the reconstructive nature of memory and the influence of social processes (Bartol & Bartol, 1998; Lloyd-Bostock & Clifford, 1983). Nevertheless, it was the pioneering work of Loftus (1979) that was to recast the way in which eyewitness research was to continue. Loftus argued that there existed a cavalier attitude towards eyewitness testimony, and both legal experts and the general public greatly underestimated the malleability of memory.

The accuracy of observer recollection clearly becomes important when placed within the investigative and legal process; and whilst it may be a difficult task to assess the veracity of a witness's account, many commentators concede eyewitness testimony is extremely persuasive. Loftus, for example, describes a case in North Carolina, USA, where on the 15th May, 1975 an assistant manager of a store was forced by two men at gunpoint into a car. He only had a brief opportunity to see their faces before they pulled stockings over their faces and laid him down in the back of the car. At arrival at the store the men demanded the contents of the safe, but following the manager's

repeated knowledge of the safe combination they left taking only $35 from his wallet. The manager in his recollections to the police included that one kidnapper looked like a man who had recently applied to the store for a job and that the vehicle involved was a white Dodge Dart.

Three days later the police arrested two brothers; the Sawyers, in a vehicle resembling a Dodge Dart. Neither had applied for a job at the store and both denied knowledge of the incident. Despite evidence that the Sawyer brothers where elsewhere at the time, the jury was prepared to accept the testimony of the victim and convict them. Only a written confession from a youth prisoner led to a later pardon. Strong testimony effects have also been evident in other cases, which have later been revealed to be not accurate as first thought (see Loftus & Ketcham, 1991).

Nevertheless, the Rand Corporation (1975) report had previously noted the main predictor of whether a crime was solved was the completeness and accuracy of eyewitness accounts, particularly in the absence of other clues. A study by Sander's (1986) further illustrated the significance of eyewitnesses in the Criminal Justice System. When Sheriff's' detectives and deputies were asked: "What is the central and most important feature of criminal investigations?" the response in the main was "eyewitnesses". Indeed, a report from the USA in 1989 estimated that around 77,000 people a year were charged with crimes solely on the basis of eyewitness evidence (Goldstein, Chance & Schneller, 1989). The continuing importance of eyewitnesses within the Criminal Justice System has been highlighted in a recent survey of 159 police officers in the UK. This revealed that 36% of officers said that

eyewitnesses 'always' or 'almost always' provided the major leads for an investigation, and that 51% said 'usually' (Kebbell & Milne, 1998).

However, while eyewitnesses are clearly important in terms of both initial and ongoing investigation procedures, in the UK, once a case has been brought to court by the Crown Prosecution Service (CPS), the evidence witnesses impart during this final stage takes precedence in terms of the informational influence carried. It is significant, therefore, that the vast bulk of the extensive literature on eyewitness testimony has concerned only those factors that may influence testimony during the initial investigatory process.

This thesis, therefore, is concerned with the examination of eyewitness memory at the final point of retrieval, that is, in court, and in particular, during the cross-examination process. However, before examining what factors are likely to influence accuracy and confidence of witnesses during cross-examination, it is first necessary to provide a background regarding the kinds of processes and procedures that are inherent in the English Legal System and occur in the courtroom.

1.2 The English Legal System and general principles of procedure and testifying in court

The general principles regarding procedures relating to testifying in court can be summarized as follows (McConville, Hodgson, Bridges & Pavolonic, 1994; Murphy, 1994; Stone, 1988). In court, every witness called is subject to examination by the party calling him or her, all other parties, and the court. Examinations consist of examination-in-chief, cross-examination, re-examination, and possible examinations by judge and/or bench, usually for

clarification purposes. The examination-in-chief is a procedure that rests upon the notion of gaining the trust of the court and jury in the witness. To this end, a witness's own counsel will encourage the witness to provide a free narrative account of events and importantly, leading and suggestive questioning styles are, by the rules of evidence, disallowed. Consequently, the examination-in-chief procedure *must* be conducted without leading questions as such questioning is regarded as improper (also disallowed in re-examination), in the sense that such questions tend to put words into the witness's mouth, or suggest directly to the witness the evidence counsel expects to receive. The rationale behind this process is that the court is concerned to hear any evidence given by the witness in the words of that witness and not those of the advocate. However, this unwritten rule is easy to state but far more difficult to apply. In an attempt to avoid leading questions, advocates are taught to apply the 'two-for-one rule': that is, a general question will be asked first inviting a yes or no answer, followed by an invitation for the witness to 'fill in' the specific detail, without leading. Nevertheless, control is still retained by counsel as to the evidence required and thus that deemed important to the case.

Memory-refreshing documents containing statements made or verified by the witness may be used while giving evidence in the witness box (e.g. a police notebook), as the giving of evidence should be a test of *accuracy* rather than one of simple memory, although previous consistent statements may not be used to enhance the witness's evidence by relating to the court what he or she has said on the same subject on another occasion.

Cross-examination, however, has a different purpose; that is to establish the creditworthiness of the witness and the weight that should be

given to evidence of that witness. Thus, with regard to cross-examination, Du Cann (1964) cites Lord Hanworth's, Master of the Rolls, statement that:

> "Cross-examination is a powerful and valuable weapon for the purpose of testing the veracity of a witness and the accuracy and completeness of his story" (p. 95).

To this end, a witness' knowledge of the facts, impartiality, truthfulness, respect for oath or affirmation and general demeanour, are challenged at this point. Witness character, bias and unreliability are also open to question. Given that, the use of previous inconsistent statements is probably the most effective way of discrediting a witness's account, it is no surprise therefore that cross-examination strategies tend to be built around the development of this consequence.

Because, on the whole, procedures inherent within cross-examination have long been thought by the legal profession to be crucial for probing of the accuracy of evidence obtained in the examination-in-chief and the exposure of unreliable or dishonest evidence (Murphy, 1994; Stone, 1988), a firm rationale has developed historically in the legal culture whereby leading questions may always be permitted during cross-examination, providing they do not invite irrelevant or inadmissible answers (R v. Thompson, 1912). Legally speaking, therefore, leading questions should be distinguished from those which are "improper; i.e. those which do not suggest the desired answer but assume the existence of a disputed fact" (Cross, Wilkins & Tapper, 1986; p. 82). The admissibility of leading questions is based upon the notion that these specific

highly suggestive forms of questioning (i.e. 'lawyerese' questions) serve to calibrate or assess the memories of witnesses. Leading questions are usually suggestive to a degree (e.g. "The car was black, wasn't it?"), while highly suggestive questions tend to be declaratives and are more powerful in their assertions (e.g. "You've seen the black car, haven't you?"). Consequently, such forms of questioning can be characterised as direct or closed lines of questioning which aim to limit the response made and elicit the preferred answer.

1.3 Conclusion

In summary, although the topic of the reliability of eyewitness testimony has received a large amount of attention from both academic researchers and legal experts, little attention has been paid to the factors that influence testimony during the final, and in many respects, the most important part of the legal process; examination in court. Research into the influences of all aspects of the examination process in court is obviously beyond the scope of a limited thesis such as this, however, the cross-examination process would seem a particularly useful place to start.

To reiterate, the cross-examination process has been viewed by many legal experts as fundamental to determining the completeness and accuracy of a witness's testimony; to this end leading and declarative questions are seen as useful ways of enabling both witnesses and observers to assess or calibrate the accuracy of their memories. However, one may legitimately ask, what actual evidence is there that the rationale upon which the cross-examination

procedure has been built is a sound one? Does cross-examination serve to calibrate witnesses' memories; or are the techniques used likely to interfere with accuracy?

Psychological literature from a number of areas is pertinent to these questions, but most particularly, general eyewitness testimony, interrogative suggestibility and police interviewing.

CHAPTER 2

COURTROOM QUESTIONING

2.1 Eyewitness memory, language and testimony

A number of findings from the general literature on eyewitness testimony suggest that cross-examination procedures might be problematic in evoking accurate testimony.

Loftus (1979) has described and categorised eyewitness memory in terms of a three-stage process. First, is the acquisition and initial storage of information concerning the event. At this stage non-linguistic factors are the main influences on memory. For example, the nature of constructive perceptual processes means that people do not record events directly or completely as lay people may actually think, and that such perceptions may therefore be inaccurate. The conditions at the time (known as event factors), such as, time exposure to stimulus (Hall, 1980; Loftus & Kallman, 1979), event complexity (Carr, Deffenbacher & Leu, 1979; Deffenbacher, Carr & Leu, 1981; Mandler, 1980), and stress (Bohannon, 1992; Clifford & Hollin, 1981; Heuer & Reisberg, 1990; Loftus & Burns, 1982; Neisser & Harsch, 1990; Peters, 1988; Yuille & Cutshall, 1986) may all influence memory accuracy. The latter may be particularly significant in that stress arousal may be related to courtroom examination styles.

The second stage is the retention or storage of information. The retention of information is generally believed to decline over time, according to Ebbinghaus's classic research into the standard forgetting paradigm (1913),

and although some have questioned the reliability of the paradigm (Bourne, Dominowski & Loftus, 1979), psychological evidence suggests that the length of retention interval is still a critical factor (Carr, *et al*, 1979; Lipton, 1977). This again may be significant given that the time lapse between the observed event, the initial police interview, and courtroom examination may often be years.

Most important, however, for the focus of the present thesis is the third stage, that of the retrieval of information. For eyewitnesses, retrieval usually involves imparting information under some form of questioning as part of an interview process. According to Loftus (1979), at this point, linguistic factors (i.e. articles, wording, bias, style, question type) play a more vital role in reshaping memories.

Before going on to examine what the literature may have to say about possible linguistic effects upon eyewitnesses, however, it is important to look in some more detail at the meaning and role of courtroom questioning.

2.2 Questioning in the courtroom

Nearly all the literature relating to a court context, whether real or fictional, recognises the importance of questions as the primary means of obtaining information from defendants, witnesses, and in fact all those who take the witness stand in a courtroom (Baldwin, 1987; Du Cann, 1964; Evans, 1995; Glissan, 1991; Hickey, 1993; Mauet & McCrimmon, 1993; Murphy, 1994; Stone, 1988). At this time, the plight of the individual could be perceived as being pitted against the arbitrary power and control of the legal system. For the witness, who is denied the opportunity to ask questions, the procedures

inherent within the process of cross-examination may constrain both the amount and type of information the witness may give as evidence. Consequently, the functions and responses in relation to questioning in the courtroom raise concern regarding the relationship between basic paradigms of justice and fairness.

For instance, in her study of question forms within magistrate's court, Harris (1984) found that different forms of questioning were specifically targeted to a preferred outcome at a certain point in the court proceedings. To reiterate, examination-in-chief will likely comprise of certain elements; the exclusion of leading questions; the incorporation of only slight suggestion; and tightly framed controlled questioning, with a slow and deliberate rate of question deliverance, perhaps in the form of a story telling narrative. The aim of such an approach is to enhance witness' creditworthiness and to raise the jury's level of trust in that witness's testimony. Conversely, cross-examination is more likely to comprise of opposite items; for example, the inclusion of leading questions, a varying rate and type of questions (possibly in no particular order), which are designed to confuse the witness; the aim is to discredit the witness in the eyes of the jury so they may come to regard the whole of the witness's testimony as disreputable.

Many questions during cross-examination contain propositions, which require completion by the witness or completion within the question structure before a response is explicitly or implicitly asked for. Such pre-supposition is commonly used in the cross-examination procedure and is usually phrased so as to elicit a response from a two-alternative forced choice (i.e. yes or no), usually in the context of 'yea' saying, thus limiting the response that can be

made and encouraging answers preferred by the questioner (Harris, 1984). For example, it has been noted that, "... all apparently and formally neutral questions are in fact marked as expecting a positive polarity answer" (Coulthard, 1981: p. 22).

Thus, certain questions inherently pre-suppose that the proposition contained within them is correct, thereby encouraging the desired minimal answer from the respondent. However, should a witness attempt to elaborate beyond this minimal response he or she is likely to be interrupted by being asked another question, or by some other response from the cross-examiner; for example, the application of negative feedback, challenging the witness and implying that the witness is mistaken or lying, or asking (redirecting) the same question again. In her study, Harris (1984) also reported that some forms of questioning are more conducive to producing the response desired by the questioner than others, and that in the court context a large number of these exist (e.g. declaratives with interrogative tags or frames, declaratives asking for confirmation, disjunctive interrogatives). Harris concludes that the power of questioning is mainly elicited through controlled questions, which require a yes/no, disjunctive or other restricted response. Furthermore, conducive form usage together with questions containing completed pre-suppositional information, serve to maintain the traditional court position: that of controlling what counts as information.

Another significant aspect of cross-examination is that there is no clear-cut separation between information gathering and accusation. Legally speaking, it is apparent that all questions can be related to one of two issues; that is; to matters of issue or information gathering, or accusations. Many

initial questions are intended to be purely informative. However, because of the court's tendency to adopt the 'unwilling paradigm' when a witness for the prosecution is being cross-examined by defence counsel, questions in this context tend to become accusations. Thus, the context becomes the main determinant as to whether questions function as accusations or a request for information, or both. Consequently, cross-examination questioning might be conceptualised as functioning in a more accusatory fashion than one that is merely searching for the truth or the facts. Accusations convey information, and questions requesting information can often contain underlying accusations. In this regard, Atkinson and Drew (1979) found that witnesses often anticipated accusations before they were made, offering defences in response to apparently neutral questions. This is probably because witnesses come to court expecting to be accused and to have their behaviour questioned. In terms of perceptions of witness credibility, such responses to neutral questions may suggest that the witness has something to hide. Therefore, the use of conducive forms of questioning which call for confirmation of prescribed information, together with accusations to which the witness cannot respond, act as a means of control over both functions by; a) restricting the introduction of 'new' information, and, b) by allowing accusations to be made while ensuring that challenges are less likely.

Hence, while questioning procedures are the accepted means of gaining evidence from a witness, the preferred answer may not necessarily be the most accurate answer, and some linguists and legal commentators have long been concerned about the possibility that lawyerese questions (i.e. questions containing leading and suppositional phrases) in court can suggest or

even compel a response (Brennan, 1995; Danet, 1980; Danet & Bogoch, 1980; Drew, 1990; Kebbell & Deprez, 1998). Indeed, Atkinson and Drew (1979) have suggested that questions in the courtroom might be: "Built in such a way as to elicit certain preferred answers, and hence create difficulties for witnesses to disagree" (p. 154).

One reason why a witness might find it difficult to disagree with such questions is that challenging a proposition contained within such a question may require greater cognitive work than simply agreeing to it. For example, witnesses know that if they disagree with a proposition they run a significant risk of further questioning on the point, and having to justify any further assertions they make. This in turn, may result in even more acquiescent and agreeable responses.

By way of illustration, a pre-supposition embedded within a question could be viewed as a different form of conditional statement. A conditional statement has no truth or falsity, only contingents. Placed in the procedure of cross-examination one could assert that witnesses who were unable to deny the antecedent/consequent (i.e. the giving of an affirmative response) might subsequently be thought less credible than if they were to reject the basis of the proposition, thereby adding weight to testimony and creditworthiness. For instance, research into pre-suppositional calculus using thematic materials has revealed that participants consistently find it difficult and confusing to choose falsification (i.e. a correct 'negative') over confirmation, even when explicitly asked to do so (Gigerenzer & Hug, 1992; Griggs & Cox, 1982; Manktelow & Over, 1990a; 1990b; 1991). Thus, it might be that witnesses who are susceptible to cross-examination questioning strategies inherently seek

preference not to engage in challenging behaviour, or feel uncomfortable doing so, and so *'choose'* the 'positive' or easier cognitive route in regards to their testimony (i.e. saying 'yes').

Nevertheless, according to many other legal commentators, and in accordance with present practice, pre-suppositional questioning is generally regarded as an important tool that counsel can employ acting as a method of verification of the facts and witness credibility (typical sources quoted are, Fillmore, 1969; Frege, 1975; Russell, 1905; Sellars, 1954). According to Walton, pre-supposition is "a proposition that one becomes committed to automatically, simply by giving any direct answer to a question" (1989, p. 31), and in general it denotes a relationship between two propositions, statements or sentences, whereby the truth or falsity of one may affect the 'truth value', 'appropriateness' or 'honest usability' of the other. Thus, within the legal arena it is generally contended that asking questions containing false pre-suppositions is a normal, useful, and effective procedure for verifying doubtful information and introducing new information (Hickey, 1993). Indeed, Hickey comments that pre-supposition; "performs a crucial function in the objectives of witness information and witness credibility" (p. 99) and thus represents " a legitimate, effective and perfectly respectable contribution to the judicial process" (p. 109).

One of the main assumptions underlying the use of pre-supposition is that it serves to expose the witness who is untrustworthy. Hence, it is argued, during cross-examination, the aim of both parties is to put forward a plausible and credible account resembling what the judge and jury (i.e. being reasonable persons) could believe. Plausibility being: "a matter of the internal coherence

of that narrative" (Jackson, 1988, p. 171). This clearly involves not only witness creditworthiness but also the credibility of the information they present. If a witness fails to give the impression of being reliable, trustworthy, and honest, (i.e. both accurate and confident) then that witness' story will carry little weight, irrespective of its criterion of truth. Cross-examination strategies, therefore, aim to break the internal coherence of a witness' testimony on the assumption that, if this is not possible, then the witness must be telling the truth.

Furthermore, even if counsel does not achieve a witness' acceptance of a pre-supposition does not necessarily indicate failure; at least the jury have heard it. Additionally then, pre-suppositions allow also for the introduction of new information, not as facts, but as possibilities. This cultural legitimacy of pre-suppositions built into questions in the courtroom is, therefore, based upon the notion that they represent a natural method of counteracting or compensating for potential weaknesses in individual witnesses, such as, poor memory, suggestibility, unpleasantness or hesitancy, lack of ability to estimate important elements, and so on.

Of course, if one assumes that cross-examination of this kind tests the credibility of the witness to the full (i.e. a witness who rejects all attempts to be led, must be accurate in what he or she says), then, from the cross-examiner's point of view, there may be a downside if the witness refuses to be led or comply with the pre-supposition. Accordingly, whilst questions that include pre-suppositional information appear commonplace during cross-examination, some legal advisors have openly asserted that asking such questions can be unwise and by doing so one can seriously risk losing a case

(Du Cann, 1964; Evans, 1995; Glissan, 1991; Mauet & McCrimmon, 1993). Consequently, Evans (1995) has argued that a number of cautionary rules should be followed during cross-examination; these are: "do not ask a question unless you know what the answer will be" (p. 141); when the witness does not give the desired response, "do not suddenly draw back with a start"; and "ride the bumps" (p. 142). And, more generally, (i) advocates should give the impression that they know what they are doing; (ii) they should be aware of the facts of the case; (iii) they should be in control; and importantly, (iv) they should not alienate the decision-makers. Thus, on the whole, Evans argues that to ask a leading question without being 'absolutely certain' about the response leaves the advocate's case open to significant uncertainty in that any unexpected response made by the witness can 'backfire' on counsel who may then find him or herself being 'lead' by the witness.

2.3 The influence of wording, bias, and question type

Despite the general contention by legal commentators that leading and pre-suppositional statements in cross-examination perform the essential role of counteracting potential weaknesses in witnesses, there is good reason for psycho-legal researchers to suggest that usage of these types of question forms might lead to increased inaccuracy in relation to witness responses.

To reiterate, the function of a question is to elicit a particular response. However, as Oppenheim (1966) has emphasized, the accuracy of the response may depend crucially on the way questions are worded and the type of question asked. Numerous research studies have suggested this to be the case.

For instance, as early as 1915, Muscio used different question forms: the 'indefinite' (Did you see *a* dog?), the 'definite' (Did you see *the* dog?), plus the 'conjecture' (*Was* there a dog?). Muscio reported that definite and indefinite articles (*the* versus *a*, respectively) produced higher false rates of responding. Loftus and Zanni (1975) also showed that the same small change in the wording of a question in a situation where participants viewed a film of an automobile accident produced very different responses. Again, the articles *'the'* (definite) and *'a'* (indefinite) were the only change made within the questions. The use of the indefinite article *'a'* produced many more 'I don't know' responses. Indeed, some psychologists have stated that if a speaker has already seen a particular item, and assumes the listener is also familiar with it, that speaker will use article 'the' (Anderson & Bower, 1973; Brown, 1973). Moreover, participants in these studies who were subsequently asked *'the'* questions reported having seen something regardless of whether it was present in the film. It can be noted, however, that Zanni and Offerman (1978) failed to replicate the latter, though they used a smaller sample of participants.

In another study, Loftus and Palmer (1974) used questions involving the words 'smashed', 'collided', 'bumped', 'contacted' or 'hit'; when the word 'smashed' was used, this consistently elicited a higher estimate of speed of the car than when other words were used. Additionally, Loftus, Altman and Gaballe (1975) looked at witnesses' later recollections of a videotape of a classroom disruption. Half of the participants were asked 'active' questions (i.e. using more aggressive language); half received 'passive' questions. One week later, in response to a series of scales about the event, participants

previously receiving the active questionnaire rated demonstrators as significantly more noisy, violent, belligerent and antagonistic.

Other researchers, such as Cady (1924), Marquis, Marshall and Oskamp (1972), and Dent (1978) found that when participants were given questions that encouraged free narration, those who gave free narratives were most accurate, though their narrations were also the least complete. Conversely, direct questions yielded less accurate but more complete responses. However, Marshall, Marquis and Oskamp (1971) found that the completeness of the testimony increased much more than accuracy decreased under all conditions of interrogation. It was also shown that in the case of content describing *person,* the increase in information was accompanied by a more serious loss in accuracy than was the case for *action, sound* or *object* content. This suggests that the effects of interrogation on the reliability of testimony are not necessarily uniform across stimulus categories.

With regard to other question types, Lipton (1977) had participants view a short murder film in an attempt to assess accuracy and completeness of testimony. Findings in percentage terms showed the following: for accuracy and completeness, respectively: questions encouraging unstructured testimony 91%, 21%, open-ended questions, 83%, 32%, leading questions, 72%, 79%, and multiple-choice questions, 56%, 75%. Thus, Lipton replicated the previous findings that free narration was most accurate but less complete, though some research has failed to confirm these findings (Clifford & Scott, 1978). In another study, Lipton (1977) used multiple-choice questions to examine the effects of positive question bias (correct answer contained), and negative bias (no correct answer). These were compared with the performance

in a neutral condition (open-ended questions). Highly significant effects were observed for both accuracy and completeness, respectively: open-ended questions, 83%, 32%, positive questions, 76%, 78%, and negative questions, 52%, 73%. On the basis of this evidence Penrod, Loftus and Winkler (1982) conclude that the evidence favours;

> "... an investigative method that initially uses free or controlled narrative and is then followed by structured questions. This method has the best chance of eliciting an initial highly accurate rendering of facts that can then be elaborated on with the structural questions" (p. 139).

Interestingly, free narration appears to be somewhat frowned upon within court examination procedures, particularly during cross-examination. This is probably because giving witnesses the opportunity to explain fully all the evidence they have at their disposal may in some way jeopardise counsel's line of enquiry (i.e. bringing into issue previous information which has already been discarded as not relevant or prejudicial to counsel's case). In other words, control would be ceded to the witness. Instead, the main form of questioning in cross-examination is pre-suppositional, leading, and sometimes confusing.

2.4 Leading, pre-suppositional and confusing questions

To reiterate, the main legal assumption made in terms of cross-examination is that it is difficult to mislead a witness who is *honest* in his or her testimony. By implication, therefore, a witness who can be misled (i.e. through certain

questioning procedures) could be regarded as dishonest, subsequently raising the issue of lying to the court, and ultimately bringing the credibility of any given testimony into question. Indeed, the use of leading and pre-suppositional questioning techniques is often promoted as the way in which the 'truth' can be attained and justice done. But is testing the veracity of the witness in this way more likely to lead to accurate responses?

The Oxford Dictionary definition of 'leading' in the context of a question is 'to so frame as to prompt the answer desired' (1967). The fact that questions can produce distorted responses in witnesses, because they are phrased in such a way as to suggest the wanted response has been known for some time (see, for example, Binet, 1900, 1905; Stern, 1910, 1938, 1939).

Doubts have been raised about the use of pre-suppositional questions by a number of more recent studies which have shown that such question types may actually impede witness accuracy (Dent, 1978; Dent & Stephenson, 1979; Kebbell & Giles, 2000; Kebbell & Johnson, 2000; Loftus, 1975; Marquis, Marshall & Oskamp, 1972, Turtle & Wells, 1988), and unwittingly provide obstructions to the truth (Perry, McAuliff, Tam, Claycomb, Dostal & Flanagan, 1995; Westcott, 1995). Thus Loftus (1975) cites four separate experiments conducted on a total of 490 participants, which showed that pre-supposition in a question could erroneously influence an answer given to a subsequent question regarding the pre-supposition. Further, she notes that;

> "The subject might treat the pre-supposed information as if it were an address, a pointer, or an instruction specifying where information related to that pre-supposition may be found" (p. 563).

Other studies have included 'lawyerese' questions that are deliberately complex and confusing. For example, Kebbell and Giles (2000) questioned participants one week after they had viewed a videotaped incident; they found that when questions were asked in a confusing lawyerese manner that included leading questions, witness accuracy and the relationship between accuracy and confidence was reduced, compared with a simple question condition. Perry, McAuliff, Tam, Claycomb, Dostal and Flanagan (1995) also found that confusing lawyerese questions impeded accuracy. These researchers compared different question types (i.e. negatives, double negative, multi-parts, difficult vocabulary, and complex syntax) with age groups ranging from 5-19 years. They found that negatives, double negatives, and multi-parts posed the greatest problems for all age groups.

However, evidence suggests that other factors may moderate some of these effects. For example, Dodd and Bradshaw (1980) replicated Loftus's effect for pre-suppositional questions, but found that the effect could be cancelled if the person being questioned perceived the questioner to be biased, and was thus suspicious of the intentions of the questioner. An implication of Dodd and Bradshaw's finding for the courtroom is that, should opposition counsel be perceived by the witness to be biased, this might place witnesses on their guard. That is, they may adopt a 'suspicious set' (Gudjonsson, 1989b; 1992) and be reluctant to accept suggestions contained within questions. Indeed, other researchers have shown that suspiciousness, whether generalised or developed in response to a given situation, results in less compliant behaviour (Milberg & Clark, 1988; Stricker, Messick & Jackson, 1967).

This suggests that alerting witnesses to possible biases may serve as a safeguard against compliant behaviour during cross-examination. This does not mean, however, that briefing will necessarily lead to more useful testimony. Wells, Ferguson and Lindsay (1981) found that witnesses who had been 'briefed' through rehearsal of their testimony; for example, giving sample questions that might be asked by a cross-examiner, and warnings that a cross-examiner will look for inconsistencies, were likely to show greater confidence regardless of accuracy. The authors suggest that;

> "witnesses seem to convince themselves of their accuracy, perhaps because the rehearsal involves a biased search for consistent supporting evidence" (p. 694).

2.5 Conclusion

In sum, courtroom cross-examination is an interview process that contains pre-suppositional question forms and leading questions, based securely on the belief that these components will calibrate witnesses' memories, and establish their credibility as data gatherers. However, there has been some disagreement in legal circles as to the efficacy of such techniques, and the psychological literature on eyewitness testimony suggests that there may be some grounds for concern in this respect.

Nevertheless, the literature also suggests that some of the negative effects of 'lawyerese' questioning might be reduced if witnesses were to be suspicious of the way in which they are interviewed.

CHAPTER 3

INTERROGATIVE SUGGESTIBILITY

The psychological literature on interrogative suggestibility may have important implications for the present thesis in two respects; 1) it may help to define some of the situations in which witnesses are most likely to produce biased testimony; and 2) it may help to define some of the characteristics of witnesses who are most likely to be adversely affected by the way they are cross-examined.

3.1 Suggestibility and interrogation

Defining the term 'suggestibility' is not easy. Marcuse (1976) argues that the term may have a multiplicity of meanings including, susceptibility to influence from another without one's consent, a tendency to accept uncritically ideas implanted in the mind consciously or unconsciously, and possessing a submissive tendency. According to Gudjonsson (1992), however, the form of suggestibility most relevant to interview performance is interrogative suggestibility. This is to be distinguished from what many have termed 'primary suggestibility', which relates more to ideo-motor and cognitive responses, such as non-volitional movements, and hallucinations, in response to suggestions that these experiences will occur (see Hull, 1933; Eysenck, 1947; Evans, 1967, 1989; Wagstaff, 1991). Interrogative suggestibility has been construed by Gudjonsson (1987) as more related to the notion of 'secondary suggestibility'; i.e. the form of suggestibility traditionally more

related to general susceptibility to social influence, encompassing concepts such as 'gullibility', and 'obedience to authority' (Binet, 1900; Eysenck, 1947; McDougall, 1908).

For example, in 1974, Milgram dramatically showed the effects of authority, status and power on obedient behaviour. He concluded that the extensive willingness of participants to uncritically obey the experimenter in his study was due to the hierarchical relationship between the experimenter and the participant, which put the participant under strong pressure to obey (see also Milgram, 1963; 1965). Such research raises an important point about the extent to which the etiquette intrinsic to a particular situation can influence human behaviour. For instance, a number of investigators have pointed out that during routine interrogation some witnesses may obey instructions that ordinarily they would resist, and that such responses could conceivably be mediated by other factors such as, a desire to be liked, an eagerness to please, a need to maintain self-esteem, and the need to fulfil role obligations and expectations, together with the possibility of affective reaction to the idea of conflict and confrontation (Irving and Hilgendorf, 1980; Watson & Friend, 1969). Also, the whole courtroom situation may be considered to contain a strong hierarchical element that may place pressure on witnesses to conform to what is expected of them. The phrase *'I promise to tell the truth, the whole truth, and nothing but the truth'*, seems to epitomize this from the beginning. Cross-examination interviewing procedures, in particular, might capitalize on the general hierarchical ambience to induce pressures upon witnesses by encouraging, even demanding, through the questioning strategies employed, that witnesses respond in particular ways.

This is significant, given that, according to Gudjonsson (1992), interrogative suggestibility is likely to be maximised during situations that include the following aspects:

1. A questioning procedure.

2. Questions which are mainly concerned with past experiences and events, recollections, and remembered states of knowledge.

3. In situations which contain a strong component of uncertainty, which is related to the cognitive processing capacity of the individual.

4. In situations that tend to be highly stressful with important consequences (Gudjonsson, 1989a).

Immediately, one can see that all of these components directly map onto the procedure of cross-examination of witnesses in the courtroom.

Nevertheless, while it is true that legal advocates have long been interested in individual witnesses' levels of suggestibility, it is only relatively recently that interrogative suggestibility has attracted interest because of its pivotal relevance to the Criminal Justice System.

3.2 Suggestibility in the courtroom

Schooler and Loftus (1986; 1993) have identified two perspectives on interrogative suggestibility; a) the 'individual differences' approach, and b) the 'experimental' approach. Typical of the first is Gudjonsson and Clark's

(1986) model of interrogative suggestibility; according to this, suggestibility is dependent upon the coping strategies people generate and implement when confronted with the uncertainty and expectations of the interrogative situation. The experimental approach, however, emphasises understanding the conditions under which leading questions are likely to affect the verbal accounts of witnesses. Within this approach, the central cognitive mechanism labelled 'discrepancy detection' (i.e. the idea that one can 'catch' the discrepancy between post-event suggestions and the original memory) is thought to mediate interrogative suggestibility (Schooler & Loftus, 1986).

However, as with suggestibility generally, not all authors agree on a definition of interrogative suggestibility. For example, Powers, Andriks and Loftus (1979) define interrogative suggestibility as; "the extent to which, people come to accept a piece of post-event information and incorporate it into their recollection" (p. 339). While this highlights the importance of memory and cognitive processing, Gudjonsson (1992) argues that, "it has not been proved that people necessarily incorporate the suggested information into their recollection" (p. 114). Hence, Gudjonsson and Clark (1986) define interrogative suggestibility as:

> "The extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning, as the result of which their subsequent behavioural response is affected" (p. 84).

This particular definition implies five interrelated components; a social interaction; a questioning procedure; a suggestive stimulus; acceptance of the stimulus; and a behavioural response. Arguably, these components map closely onto the context and procedures of cross-examination.

Another concept identified by Gudjonsson & Clark (1986) as one that might induce suggestible responses, if accepted by the interviewee, is negative feedback. They define negative feedback as "a signal communicated by an interrogator to a witness, after he or she has responded to a question or a series of questions, intended to strengthen or modify subsequent responses of the witness" (Gudjonsson & Clark, 1986; pp. 93-94). This also has obvious implications for cross-examination. In the courtroom, negative feedback occurs when the same question is repeated or redirected because the answer given is not acceptable to the examiner. This might lead the witness to adapt subsequent responses to accord with the expectations reflected in the examiner's style of questioning. It is also possible that, while a witness' initial memory might be good, the negative feedback questioning style may bring doubt into the minds of jurors, influencing the overall credibility rating of that witness.

Nonetheless, as yet, the actual types of negative feedback that might be applied during examination in the courtroom have not yet been differentiated and investigated empirically. For example, one could classify negative feedback styles as 'overt' and 'subtle'. In this respect, 'overt' forms of feedback might read; 'think about this again', 'try and be more accurate', or perhaps, 'consider your answer carefully'. A more covert and 'subtle'

approach, however, might consist of redirecting questions such as; "is it a possibility that you might be mistaken, and that....?", and so on.

In the courtroom, negative feedback is most commonly used during cross-examination when the examiner attempts to challenge directly a specific memory of the witness and to change his or her answers. This is important, as Schooler and Loftus (1986) argue that "negative feedback may reduce witnesses' confidence in their own memories thus making them more susceptible to leading questions" (pp. 107-108). According to Schooler and Loftus this may occur because negative feedback can reduce the ability to detect discrepancies, and "recollections are most likely to change if a person does not immediately detect discrepancies between post-event suggestions and memory for the original event"; i.e. fails at discrepancy detection.

However, Schooler and Loftus do not distinguish between the possible differential effects of types of feedback. For instance, more generally, they argue that discrepancy detection is influenced by two main factors, a) "the strength of the original information in memory" and, b) "the manner in which the post-event suggestion is influenced" (p. 108). With regard to the first, Hertel, Cosden and Johnson (1980), indicated that participants are more likely to incorporate misleading suggestions into their recollections when there is a long interval between viewing the event and the presentation of post-event suggestions. As mentioned earlier, the relevance here in relation to a witness giving evidence in court is that the amount of time it takes for a criminal case to be brought to court can, at best, be months, and at worst, years. Thus, Hertel *et al.* suggest that post-event suggestions are least likely to impair

discrepancy detection when encountered very close to viewing the original event.

However, by varying the sentence construction of misleading suggestions, Loftus (1981) found that explicitly directing participants' attention to the misleading information made them more willing or able to scrutinise their memories and detect discrepancies. In another study, Greene, Flynn and Loftus (1982) advised participants, prior to a reading task, to be on the lookout for misleading information. They found similar results. Additionally, Tousignant, Hall and Loftus (1986) found that just asking participants to read post-event narratives slowly increased discrepancy detection. Arguably, therefore, overt feedback that forces the witness to reconsider his or her memory might actually have a positive effect.

Conversely, Tousignant, *et al.* (1986) have argued that the uncertainty induced by the feedback might "facilitate suggestibility by reducing the likelihood that a witness will experience a discrepancy between the original event and the subsequent suggestion" (p. 107). Further, possible increases in anxiety due to such uncertainty may also decrease the witness' ability to scrutinise the question content, and "presumably people who tend to be less certain as a result of poor memory abilities are less able to catch discrepancies between the original event and subsequent suggestions" (Tousignant, *et al.*; p. 110). Thus, there may also be a tendency for people to be most influenced by 'unmemorable suggestions' (such as peripheral details of a crime scene or event). Notably, in cross-examination, it is usual for counsellors to target topics for cross-examination about which the witness appears to be uncertain (Baldwin, 1987) i.e. those that most likely are difficult to remember. From the

cross-examining lawyers' point of view, the strategy of targeting difficult items makes sense, as cross-examiners would be less likely to influence the witness' response to an item that is easy to remember (such as the sex of an attacker).

Another implication of the fact that lawyerese questioning styles and strategies might exacerbate levels of uncertainty, is that the uncertainty, and generally 'being made to look as though one is wrong', might result in stress; and there is a large literature indicating that stress may affect witness' accuracy.

3.3 Eyewitnesses, stress and performance

The Yerkes-Dodson Law (1908) is often referred to in relation to the relationship between arousal and performance. In the context of eyewitnesses' accounts the law has predominantly been applied in the context that a stressful event may result in inaccurate recall of an event and/or details pertaining to that event. The relationship is alleged to be curvilinear; that is, recall performance will be poor when stress is very low; and be facilitated by a moderate amount of stress up to a critical point, beyond which performance will decline. However, the impact of the stressful event may be dependent upon a witness's reaction to that event. Thus, stress influence may be a function of the interaction between the characteristics of the event and the characteristics of the individual.

Most research into the effects of stress on eyewitnesses has looked at the effects at encoding rather than retrieval. However, much of this might, nevertheless, be relevant in the courtroom.

In a survey of 63 experts on eyewitness testimony, 79% of the experts agreed with the statement: "Very high levels of stress impair the accuracy of eyewitness testimony", and 71% said that the statement was sufficiently reliable to offer it in court (Kassin, Ellsworth & Smith, 1989). However, the evidence is mixed. While some researchers have reported that negative emotional events (i.e. stressful stimuli) can be poorly retained (Clifford & Scott, 1978; Clifford & Hollin, 1981; Loftus & Burns, 1982; Neisser & Harsch, 1990; Peters, 1988), others have reported that such stimuli are well retained (Bohannon, 1992; Heuer & Reisberg, 1990; Yuille & Cutshall, 1986, 1989).

A number of studies have shown that crime seriousness can negatively affect the quality of subsequent eyewitness accounts. For example, Clifford and Hollin (1981) showed a violent interaction between a man and a woman and a non-violent interaction between a man and a woman. Participants who saw the violent film were significantly less accurate and less complete in their account than those who observed the non-violent event. Similarly, Loftus and Burns (1982), also using violent and non-violent endings, found that those who observed the violent ending were less accurate and less complete in their memory for that event. Additionally, memory for events two minutes prior to the event were also affected, perhaps implicating some disruption of normal memory processes. However, the seriousness of an event may not be all that easy to define or distinguish.

To reiterate, some studies have found that negative emotional events are well retained. Christianson (1984) reported that participants who had viewed an emotional middle section of a slide sequence recalled the main

features and the gist better than those who observed a neutral version. Christianson and Loftus (1987) also found that participants who observed emotional pictures could recall central features better, although they were less able to recognise specific pictures they had seen. This pattern remained stable over various retention intervals: at 20 minutes, at 2 weeks, and 6 months later.

Importantly, however, a review carried out by Christianson (1992) distinguishes between witness arousal (i.e. stress per se) and emotionally arousing events (i.e. the valence of stressful stimuli independent of the arousal of the witness). Thus, differentiating between situations in which the 'to-be-remembered' event is accompanied by emotional arousal that is evoked by the 'to-be-remembered' material proper, and situations in which the source of the arousal is dissociated from the 'to-be-remembered' event. As mentioned previously, generally speaking, assumptions made in respect of emotional stress and memory have been done on the basis of recollection of those events which are also emotionally arousing and stressful, thereby not taking into consideration the possible complexities of stress. Numerous interaction effects have been shown between several factors, including; the type of event (stress invoking or not) and type of information (central or peripheral; Burke, Heuer & Reisberg, 1992; Heath & Erickson, 1998; Wright & Stroud, 1998), type of test (free recall, cued recall, recognition Christianson & Nilsson, 1984; Davis, 1990; and Wagenaar, 1986; respectively), or time of test (immediate or delayed; Burke, Heuer & Reisberg, 1992; Christianson, 1984; respectively). Similar complexities have been shown in relation to weapon focus: i.e. a phenomenon linked to stress and violence, and central and peripheral detail, where witnesses, arguably, might focus on a gun or knife at the expense of

other event information (Goodman, Hepps & Reed, 1986; Loftus, Loftus & Messo, 1987; Maass & Kohnken, 1989; Mitchell, Lovosky & Mather, 1998; Tooley, Brigham, Maass, & Bothwell, 1987). As a result, Christianson (1992) concludes that "the way emotion and memory interact is a very complex matter" (p. 303), and that the belief that high emotional arousal leads to less processing capacity is an "overly simplistic" one (p. 284). Indeed, a more recent survey of experts carried out by Kassin, Tubb, Hosch, and Memon (2001) found that a strong consensus rate of 80% had not been reached on this issue. This indicates that eyewitness testimony research in relation to stress is not presently sufficiently reliable enough to present in court. For a more detailed analysis and discussion of the highly complex relationships between emotion and autobiographical memory see Read (2001).

Nevertheless, on the basis of some of these findings, it is possible that the different styles of court examination interview could differentially affect witness' accuracy through the stress produced; though in which direction has yet to be established. Additionally, such questioning might also impact upon jurors in the sense that they might differentially attribute witness accuracy and confidence according to the style of questioning (for example, they might perceive a stressed witness to be more inaccurate or less confident).

3.4 Conclusion

In summary, the literature on interrogative suggestibility indicates that procedures related to cross-examination in court may influence the accuracy of the witness through the moderating effects of suggestibility and/or abilities to detect discrepancies. That is to say, they involve a questioning procedure

where the questions relate to memories of past experiences; contains a strong element of uncertainty which is related to the cognitive processing capacity of the individual, and in a situation that tend to be highly stressful with important consequences. Thus, under these conditions, some witnesses might respond in ways that they would not ordinarily do so for a variety of reasons that might affect levels of confidence and accuracy in those witnesses.

However, although these possibilities have yet to be explored in the courtroom, the notion that interrogative interviewing procedures might affect witnesses' memories is not a new one. For example, the police interview was subject to inquiry some 20 years ago surrounding concerns that some of its procedures might be inherently problematic for witnesses (Wagstaff, 1982). It is possible therefore, that police and psychologists have investigated problems that could have potential relevance to questioning styles used in the courtroom.

CHAPTER 4

POLICE INTERVIEWING

The testimony of an eyewitness ordinarily involves retrieval of information under some form of questioning as part of a police interview process. In fact, interviews carried out by the police are one of the most important methods for finding out relevant facts related to an enquiry. Hence, police interviews have been 'flagged' during the last decade, or so, as the 'conversational core' of policing. The way in which the police conduct witness interviews, therefore, remains a matter of public policy and concern (Shepherd, 1988; 1993).

Up until recently the police service did not train officers in carrying out skilled interviews, and policemen and women generally acquired their skills largely from observation. However, this method of observation training tended to engender a cultural regeneration of practice that did not fulfil the aims and objectives of the police interview: that is, to obtain accurate, relevant and complete accounts from the interviewee. It can be noted that lawyer courtroom practice training (i.e. pupilage) also consists largely of lengthy observations of a mentor; with possibly the same consequences.

More recently, however, both psychologists and the police have conducted research to examine the nature of evidence obtained through the standard police interview. An examination of this literature may highlight similarities between and problems associated with standard police interrogative practices and the tactics used during cross-examination in court.

4.1 The standard police interview

In 1987, Fisher, Geiselman and Raymond provided the most comprehensive critical description and review of the standard police interview to date. In their review, Fisher *et al.* noted that police interviews commonly included a very loose structure with the majority of questions asked in a direct way, and little if any support was given to help eyewitnesses to remember and/or enhance their memories. Indeed, a notable principle of memory is that much more information is held than can be retrieved at any one time, thus a failure to retrieve and so activate information contained in memories, is probably one of the major difficulties for eyewitnesses.

Fisher and colleagues proposed that the provision of retrieval cues might tap important information stored in memory, and they highlighted techniques used by the police that might hinder this important retrieval process. Consistently occurring problems in police interviewing included the following; negative phrasing, non-neutral wording, inappropriate language, staccato style questioning, distractions, judgemental comments, together with the lack of identification of potential leads. However, the most persistent difficulties were identified as *interruptions*, an *excessive use of question-answer* format, and *inappropriate sequencing* of questions.

For instance, in Fisher *et al.'s* work, witnesses were on average interrupted 11 times, and typically only 7.5 seconds after they had begun to respond. Interruptions affect breaks in concentration, thus requiring a constant switching of attention, and possibly inadvertently setting up expectations in the witness to be more passive during the course of the interview. This could lead to less information being generated, particularly during open-ended

questions (i.e. free recall), which tend to produce more accurate information (Yuille & Cutshall, 1986). Interestingly, the idea that a constant switching of attention might be required to deal with interruptions again highlights the possibility that individual cognitive processing (i.e. switching) capacity might be important. Perhaps, some measure of such cognitive capacity might predict individual witnesses' susceptibility to lawyerese questioning during cross-examination. Further, it could be conceived that the inclusion of negative feedback within a cross-examination procedure might act as an interrupting strategy.

Additionally, and typically, most police interviews rely on the question-answer format as information is produced quickly. However, police interviews rarely base interview techniques around open questions. For example, a recent study by Davies, Westcott and Horan (2000) found that investigative interviews conducted by police officers with children used open-ended questions only 2% of the time. Indeed, most questions used are closed (e.g. 'What colour was the attacker's hair?'), and while these are of practical use in that they aid in directing the witness to important information, a major difficulty is that often little time is given for the witness to engage in a more focused type of retrieval that might lead to the retrieval of other important investigative information. Basically, if the witness becomes disengaged and remains passive then it will be "virtually impossible for the witness to retrieve detailed events from memory" (Fisher *et al.* 1987; p. 181). This excessive use of closed question-answer formats, in addition to a lack of focused retrieval, also has relevance to the cross-examination procedure in court. As stated previously, cross-examination includes styles of questioning that could be

classed as closed and leading. Hence, these procedural aspects of cross-examination might encourage passivity in the witness, possibly minimising their abilities to retrieve important information, and/or detect important discrepancies.

In relation to inappropriate sequencing of questions, similar problems arise. For example, the 'dropping' in of general knowledge or opinion-based questions during the police interview (e.g. 'Why do you think that happened?') might also result in the constant shifting of retrieval attention. During cross-examination, while counsellors are encouraged to 'tell a story to the jury' cross-examination is not a reflection of the in-chief examination. A cross-examination can, therefore, be incoherent in its structure; i.e. the cross-examiner can test evidential points made during the in-chief examination at any time during the interview.

Further, when a witness is asked a lawyerese question that contains potentially discrepant material, this requires the witness to; a) listen carefully to the question, b) comprehend the question, c) check the content of the question material with that in memory, d) formulate an answer, and e) articulate a response. This suggests that cross-examination questioning might require high levels of cognitive ability in witnesses; a factor which may be particularly relevant given that questioning that requires shifts in retrieval attention can result in a 19% reduction in witnesses' performance (Fisher & Price-Rouch, 1986).

Of course, it has been recognised that the 'ideal' police interview may not necessarily be practicable, particularly as other factors, such as those relating to the eyewitness (i.e. poor verbal and/or cognitive skills, personality

and/or intoxication, trauma and/or fear), and the logistics of the situation, together with the requirements of the investigation, may also detract from what is needed in terms of effective retrieval of information from memory. Essentially, however, according to Fisher, Geiselman & Raymond, the aim of the effective interviewer should still be "to determine and construct the cue, to unlock the fact, but not *lead* the witness" (1987; p. 179). The emphasis on not leading the witness has obvious relevance to the cross-examination interview, particularly given that *leading* questions are always permitted.

At this point, it is interesting to note that some psychological commentators have noted that "many police officers will take the view that there are good witnesses and bad witnesses, and that any difference in the quality of their statements can be put down to the individual witness rather than the interviewing technique used" (Ainsworth, 1995, p. 20). Whilst this may suggest that many police officers tend to blame the witness rather than their techniques for any failure in their interviewing, one must not dismiss the possibility that good and poor witnesses might exist in a normal population. For instance, cognitive retrieval capacities might mediate susceptibility to procedural components contained within cross-examination.

Nevertheless, regardless of a witness's inherent abilities, the focus on problems inherent in the standard police interview have since lead to attempts to maximise the accuracy of witnesses' memory retrieval performance. To this end, the Cognitive Interview (CI) was born.

4.2 The Cognitive Interview (CI)

One of the first versions of what is now known as the CI, was devised by Wagstaff (1982). However, it was Fisher and Geiselman, and their colleagues (Fisher, 1995; Fisher & Cutler, 1996; Fisher & Geiselman, 1992; Fisher, Geiselman, & Amador, 1989; Fisher, Geiselman & Raymond, 1987; Fisher, Geiselman, Raymond, Jurkevich, & Warhaftig, 1987; Fisher, & Price-Roush, 1986; Geiselman, 1996; Geiselman, Fisher, Cohen, Holland, & Surtes, 1986; Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian, & Prosk, 1984; Geiselman, Fisher, MacKinnon, & Holland, 1985; Geiselman, Fisher, MacKinnon, & Holland, 1986), who developed the idea such that now police officers all over the world are familiar with its advantages. The major areas upon which the CI concentrates are memory and communication; in that the witness needs to be able to retrieve accurate memories of the event and then communicate this again accurately to the interviewer. Fisher and Geiselman (1992) proposed that this could be better achieved by incorporating into the interview four main techniques that might enhance the retrieval process. These can be summarised as follows: 1) *Recreating the context* (i.e. get the witness to think back to the occasion when the information could be recalled); 2) *Focus concentration* (i.e. this aids the retrieval of 'image' coding; that is detailed information, in conjunction with various sensory modalities such as visual, auditory, tactile, rather than 'concept' coding; general impressions); 3) *Multiple*, and; 4) *Varied retrieval attempts*. The original CI procedure thus included four principles derived from the empirical literature on information retrieval from memory (Bower, 1967, Tulving, 1974) that were assumed to increase recall accuracy without increasing the amount of inaccurate

information remembered. The main mnemonic instructions in the original CI, therefore, are: 1) *Reinstate the context*, that is, the conditions under which the event in question was encoded (this requires witness to think about what the surrounding environment was like at the scene, to think about how they were feeling at the time and their reactions); 2) *Report everything*, however trivial it may seem (witnesses are told not to edit their accounts just because they may feel the information is not important); 3) *Recount* the event *in different orders*, that is, it seems natural to recall events from the beginning to the end, but witnesses are asked to start their recall at the end and work back, or to begin with something in the incident that inspired them most, and then work both forwards and back from that point; and 4) *Recount* the event *from different perspectives*, this includes asking witnesses to try to see the incident from the point of view of some other person present, or the interviewer asking about a particular point in different ways to maximise retrieval paths.

In continuance of their research, Fisher and colleagues (Fisher, Geiselman & Raymond, 1987; Fisher, Geiselman & Amador, 1989), recommended enhancing the CI by adding the principle of *motivating witnesses to concentrate on their retrieval attempts*. It was envisaged that this concentration would most likely be facilitated by ensuring that: 1) *Rapport* is established with the witness, such that the witness is made to feel relaxed and comfortable, with the knowledge that they will not be interrupted, that is, there should be no obvious distractions; 2) Witnesses should be encouraged to *focus* their *attention* on internal images of the episode, and be assured that he or she has unlimited time to search through their memory. Therefore, attempts are made to tap the 'image' rather than the 'concept' code. Thus, the interviewer

needs to structure the interview such that "it is compatible with the eyewitness's mental representation of the crime" (Fisher, Geiselman & Raymond, 1987; p.179). It is important then that the interviewer be careful to ensure that the witness selects the content areas for focusing; and 3) The interviewer helps to produce *'focused retrieval'* by encouraging the eyewitness to concentrate on his or her retrieval attempts, principally through motivation.

One overall effect of these techniques is that the eyewitness tends to guide, direct and select the content rather than the interviewer, hence minimising constant switching of attention.

4.2.1 Evaluation of the Cognitive Interview (CI)

The CI has attracted much interest from those in the field of forensic interviewing (for reviews see Bekerian & Dennett, 1993; Fisher, 1995; Fisher & Geiselman, 1992; Memon & Bull, 1991; Memon & Kohnken, 1992; Kohnken, Milne, Memon & Bull, 1994; Memon & Stevenage, 1996).

Fisher and Geiselman (1992) claimed that the CI "increased substantially the amount of information gathered in many different settings", and that "it worked with student and non-student witnesses; novice and experienced investigators; criminal and civil investigations; in the laboratory and field" (p. 5). In support, a number of studies have shown that the CI can produce more information than other interview techniques (Geiselman, 1996; Geiselman, Fisher, Cohen, Holland, & Surtes, 1986; Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian & Prosk, 1984; Kohnken, Thurer & Zoberbier, 1994). For example, Geiselman *et al.* (1984) compared the CI with

the 'hypnotic interview' (a controversial method of eliciting 'hypnotically refreshed' witness memories; Orne, Soskes, Dinges & Orne, 1984) and the 'standard police interview'. In this study, student participants were shown a video of an armed robbery, and then asked a number of questions about the incident. Both the CI and hypnotic interview were found to increase accurate information by 35% when compared with the police interview, without increasing inaccurate and fabricated information. Whether 'hypnosis' should be used to enhance interviewing procedures has been assessed by Wagstaff (1999), who comments; "the evidence suggests that any advantages to be gained from hypnosis over standard interviewing procedures in terms of rapport and memory facilitation mnemonics, such as context reinstatement, can be found with alternative memory enhancement techniques such as the cognitive interview" (p. 171).

In another study, Aschermann, Mantwill and Kohnken (1991) presented student participants with a short film; they were tested two to nine days later. The CI showed a significant increase in correct information recalled when compared with a standard interview condition. This was particularly notable with the initial open-ended question, although a trend also indicated that more incorrect information was also produced by the CI procedure, especially with open-ended questions. The CI has also been shown to significantly reduce the impact of misleading questions on witness accuracy (Geiselman, Fisher, Cohen, Holland, & Surtes, 1986; Geiselman, Fisher, MacKinnon & Holland, 1986).

A meta-analytic review by Koehnken, Milne, Memon and Bull (1992) of 25 'cognitive interview' studies totalling over 1,200 participants, reported

that the overall findings lent support to the CI as a superior interview technique, with all the investigations demonstrating the effectiveness of the CI in eliciting more information when compared to a standard interview. The results also found that the CI is more effective with active rather than passive observers, possibly suggesting that better results might be obtained from real witnesses. However, overall a slight increase in the amount of incorrect information was also elicited.

The CI in its 'enhanced' form has also received empirical support (Fisher, Geiselman, Raymond, Jurkevitch and Warhaftig, 1987). In this study, participants were shown a video of a violent crime and asked to recall their memory of the incident two days later. This was tape recorded and later transcribed. Fisher *et al.* (1987) found that including the additional principles outlined earlier produced 45% more accurate relevant information in police detectives' interviews of crime witnesses without increasing inaccurate recall, when compared with the original version of the CI. George (1991) also carried out an investigation into the effectiveness of the 'enhanced' CI. Fifteen student participants witnessed a staged incident during a lecture. Two weeks later police officers interviewed them using one of three techniques: a) the 'enhanced' CI; b) conversation management (CM; a procedure designed to aid social and communication skills of interviewers opening up channels of communication to elicit facts); and c) a standard police interview. A non-significant trend was observed for the CI to produce more information, with no evidence of increased errors or confabulation. Additionally, Clifford and George (1996) examined the technique in a field study where 28 policemen and women interviewed real victims and witnesses using three methods of

investigative interviewing (CI, CM and CI + CM). Again, the CI was more effective and demonstrated its advantage over other forms of interviewing, leading Clifford and George to conclude that the CI is an ecologically valid investigative technique.

While, in the main, many such studies support the efficacy of the 'original' and 'enhanced' CI, showing similar increases in relevant information and no real increases in errors or confabulations, other researchers have not reported the same (Beckerian & Dennett, 1994). For example, Beckerian, Dennett, Reeder, Sloper, Saunders and Evans (1994) reported that the CI produced significantly more incorrect information than when a standard procedure was used. Thirty-seven student participants observed a film and those in the CI condition were asked to recall a day later: a) as much as they could; b) recall information in any order; c) use context reinstatement; and d) recall everything. Those in the standard condition were instructed with a, and b. While the CI group did produce more accurate information they also reported more inaccurate information.

Discrepancies between results have raised concerns in commentators who have begun to discuss the limitations of the CI. Memon, Bull and Smith (1995) carried out a study where participants witnessed a staged armed robbery and were interviewed by officers trained in 'cognitive' or 'structured' interview techniques. No differences in correct or incorrect information elicited was found, and thus they failed to find superior support for the CI. Memon *et al.* (1995) highlighted the difficulties in training experienced police officers to use the technique. Indeed, researchers have found that, even after training, individual police officers do not perform equally well when using the

procedure (Fisher, Geiselman & Amador, 1989), probably because they do not use the interview in the way described by Fisher and Geiselman (1992).

In this respect, there does appear to be a number of operational problems in relation to the application of the CI in real-time settings. The time required to administer the procedure is one such difficulty. While Gwyer and Clifford (1997) found that enhanced CI training did produce significantly greater correct and total recall, but no greater error, than a suitably trained standard interview comparison group, discussions with police officers trained in the cognitive interview suggest that, in their experience, they often do not have sufficient time to conduct a cognitive interview (Croft, 1995; see also Fisher, Geiselman & Amador, 1989). Further, many police officers state that they prefer to use deliberate strategies (i.e. *interrupting* the witness in order to focus their attention and the use of *question-answer* as a quick method of obtaining the required information) that are specifically aimed at limiting an eyewitness's report to the minimum information deemed important (Kebbell & Wagstaff, 1996). The use of such strategies are also rationalised in a similar way by counsellors who cross-examine (i.e. that leading, pre-suppositional, and sometimes, negative feedback styles of questioning aid to serve witnesses in calibrating their own memories; Stone, 1988).

In an attempt to account for differential CI outcomes, a componential study by Boon and Noon (1994) found that not all the four main techniques employed during the CI actually increased witness accuracy significantly. Consequently, there is some confusion in the literature with regard to the relative efficacy of the CI. However, as Kebbell & Wagstaff (1996) point out, a major problem with evaluating the CI in practice is that it no longer

represents a unitary procedure but reflect an amorphous collection of interview techniques that use *all or some* of the original CI components. They conclude that any interview that attempts to overcome what are seen to be deficiencies in 'standard' interviewing could now be termed as 'cognitive'.

In sum, many of the negative findings that have been attributed to the CI may result from inadequate application of the procedures. Consequently, notwithstanding some negative findings, there seems to be a broad consensus that the procedures outlined by Fisher and Geiselman may still help to overcome some of the problems inherent in the standard police interview.

4.3 Conclusion

An examination of the problems intrinsic to, and attempts to improve, the standard police interview shows a number of issues relevant to this thesis. Clearly, some overlap exists between the cross-examination interview and the police interview in the following respects: 1) the excessive use of question and answer; 2) the use of interrupting strategies; and, 3) incoherent question sequencing. However, *none* of the strategies that make up the CI is overtly apparent in the cross-examination procedure (i.e. *reinstating context, report everything, recount events in different orders, change perspectives and establish rapport*). In fact, the nature of the adversarial system inevitably means that a cross-examiner would not attempt whole-hearted rapport with an opposition witness. In other words, if work on police interviewing is anything to go by, cross-examination procedures are more likely to militate against the production of accurate information rather than improve it. Given this, it would seem essential to examine empirically the notion upon which cross-

examination rests; i.e. that such practices are useful and effective strategies supporting memory calibration of witnesses.

CHAPTER 5

WITNESS CREDIBILITY: CONFIDENCE AND ACCURACY

5.1 Court representatives and witness credibility

Cases in which the primary facts are undisputed, and the only question is what inferences should be drawn from them, are relatively infrequent in the criminal courts. Usually, the decision as to the facts depends, in the last resort, on which of the witnesses the court representatives believe to be telling the truth. However, the assessment of a witness' credibility does not merely involve making a decision as to whether the witness is telling lies. Strictly speaking, a lie is a statement made which the speaker does not believe to be true (i.e. an intentional false statement). However, many witnesses make false statements in the firm belief that they are telling the truth.

Factors that have been considered important in determining the credibility of a witness in court include; the extent to which the witness can be ridiculed (a witness who is made to look a fool appears less credible), the character of cross-examination (i.e. a dramatic examination might increase credit given to counsel), and the demeanour of witness (e.g. any display of embarrassment and/or discomfort in the box is likely to be interpreted as lying)(O'Barr, 1982). Indeed, many judges and jury members believe that a witness who has shifty eyes, or is nervous, or stumbles in his or her speech, is certainly lying (Millar & Burgoon, 1982). Furthermore, one study has shown that most people fully believe that they can detect when someone is lying (Yarmey & Jones, 1983).

The Duke University Law and Language Project examined the hypothesis that judgements of whether a person is telling the truth are heavily influenced by linguistic variables. During this project, actual tape recordings of lawyer and witness dialogues were re-recorded by actors, with different versions illustrating different linguistic variables. Participants were then asked questions based on what they heard, and asked to rate witnesses on qualities such as competence, trustworthiness, and attractiveness; and also to rate lawyers on qualities such as intelligence, fairness, skilfulness, and control. Results showed that speakers, regardless of sex, were judged to be more trustworthy when they used what the authors termed 'powerful' as opposed to 'powerless' speech. Powerless speech is thought to approximate to language stereotypically used by white Western women (i.e. tentative, deferential, polite, exaggerated, indirect, and/or emotional). A white Western male's speech is regarded as more direct, assertive, straightforward, and rational (Erikson, Lind, Johnson, & O'Barr, 1978). These data appear to support the basic idea that more assertive and confident witnesses are viewed as more trustworthy, and, therefore, more accurate. The authors also found that narrative rather than fragmented testimony received a more favourable response, presumably because jurors inferred that lawyers would only cede some control to those witnesses who commanded some respect.

Other factors that court officials are likely to take into account when assessing witness credibility are: a) the inherent consistency of the evidence given (i.e. being non-suggestible); b) consistency with other witnesses (i.e. corroboration, where possible); c) the 'credit' of the witness (i.e. performance in witness box, physical or mental defect rendering evidence unreliable,

evidence of bias); d) observation of the witness (i.e. physical manifestation of truthfulness or mendacity such as confidence, hearing, eyesight, capacity to judge distance and/or height); and, e) the inherent probability or improbability of the evidence (Cross, Wilkins & Tapper, 1986; May, 1990; Murphy, 1994).

In relation to points a, b, and c, it is normal that counsel during cross-examination will spend much time trying to obtain from the witness some contradiction either of the witness' evidence; or of some established fact, or natural law. Crucially, whether this particular part of evidence, fact, or law, has relevant bearing on the case is actually immaterial, since the cross-examiner's objective is to use any contradiction to discredit other evidence the witness has given with regard to matters that are relevant. In respect of d and e, in countries which follow the English common law, it is considered essential to subject the witness to cross-examination, in order to 'test the credit of the witness'; i.e. to attempt to elicit answers that will enable court representatives to decide whether evidence is reliable. At this point in proceedings, cross-examination lawyers are noted to act on a number of assumptions with regard to human behaviour. One is that everyone is very capable of lying under oath; for example, the eminent lawyer, Eggleston (1978), has commented:

> "It has been my experience that honest witnesses, that is to say; witnesses who would not be prepared to lie to gain a personal advantage, will nevertheless lie to protect their good name or the good name of their friends and relations if the questions asked do not seem to them to have any bearing on the case ..."

"Few people will tell the whole truth, and many people will assert untruth, if they regard the suppression of the truth as more important than that the court should reach a correct decision in the case in which they are called ..." and furthermore,

"If they think that the question is completely irrelevant, as most cross-examination is, to the facts in issue, they will certainly withhold the whole truth, and probably lie outright ..." (pp. 158-159).

Generally speaking, however, judges are perceived to overrate the propensity of witnesses to tell the truth, as they believe that the witnesses will be as overawed as those who impose the sanction of an oath think they ought to be (Eggleston, 1978). Further, there is evidence in the psychological literature to suggest that jurors are also more likely to overestimate the credibility of the eyewitnesses (see Loftus, 1979; Loftus & Ketcham, 1991). Hence, it seems both judges and jurors tend to proceed on the assumption that witnesses are probably truthful. However, judges also assume that witnesses are either *wholly truthful* or *wholly mendacious* (Eggleston, 1978). Consequently, in practice, there does not appear to be a continuum upon which given parts of evidence might be mendacious (knowingly or unknowingly) and others truthful.

5.2 'Over-belief' of witnesses

The fact that both judges and jurors proceed on the assumption that witnesses are probably truthful is part of a general tendency amongst court representatives to uncritically accept eyewitness testimony: i.e. exhibit what has been termed the 'over-belief' effect. This in turn may have implications for verdicts; as Kennedy and Haygood (1992) state;

> "Common sense tells us that introducing the testimony of a credible witness by the prosecution should increase the probability that a jury will find a defendant guilty. Conversely, if the defence attorney can discredit the testimony of the eyewitness, the probability of a guilty verdict should be no higher than that found with no eyewitness at all" (p. 70).

Loftus (1974) used three evidence conditions to examine the notion of over-belief; 1) physical or circumstantial evidence, 2) testimony of an eyewitness, and 3) discredited eyewitness testimony. She found that the percentage of guilty verdicts increased from 18% with no eyewitness, to 72% with a credible prosecution witness. However, contrary to the commonsense argument, in the discredited prosecution condition, 68% of participant jurors still returned a guilty verdict; such results have since become known as the 'discrediting failure effect' (Saunders, Vidmar & Hewitt, 1983; p. 58). Loftus (1979) interpreted these outcomes as supporting the position that jurors place far too much emphasis on eyewitness testimony and "rarely regard eyewitness testimony with any degree of scepticism" (p. 197). It seems that eyewitness

testimony may have a particularly strong impact on observers in that it tends to be vivid, carrying distinctive information which makes it easy to remember (Nisbett & Ross, 1980), and difficult to erase from memory (Myers, 1996).

However, despite this apparent consensus, Weinberg and Baron (1982), in a replication of Loftus' work, have reported a substantial discrediting effect. Also, Saunders and colleagues (1983) noted that Loftus' trial summaries did not include judicial instructions. When Saunders *et al.* repeated the study stressing the instruction; "it is dangerous to convict solely on the basis of the uncorroborated evidence of the identification witness, as this type of evidence is potentially unreliable" (p. 62), a significant discrediting effect was found. Indeed, McCloskey, Egeth, Webb, Washburn & McKenna (1981), having found a strong discrediting effect, suggested that Loftus (1974) obtained high guilty verdicts because the trial transcripts lacked structural and conceptual verisimilitude (i.e. an air of being true; a semblance of actuality) and were not ecologically valid. For example, they lacked prosecution and defence opening arguments, cross-examination, and judges' instructions pertaining to the criterion 'beyond reasonable doubt'.

Consequently, Kennedy and Haygood (1992) concluded, "the evidence that jurors are prone to 'over-belief' is beset by conflicting evidence" (p. 73). In order to identify the sources of discrepancy they conducted a series of three experiments. These explored trial summary characteristics that may account for the discrepancies, such as the inclusion of the 'reasonable doubt' criterion, or reaffirmation of eyewitness testimony following discrediting. All conditions revealed a strong discrediting effect. They also found the following: reaffirmation of the accuracy of testimony appeared to restore

witness credibility; gender revealed no significant differences; participants were reasonably confident about their decisions; participants voting guilty made no mention of the 'reasonable doubt' criterion, whilst those voting not guilty cited the standard 25% of the time. Thus, the inclusion of the 'reasonable doubt' criterion appears to be an important aspect of the decisional process influencing the verdict reached. Further, in respect of the apparent restoration of witness credibility, Havatny and Strack (1980) refer to a 'boomerang effect' based on equity theory; thus "jurors exposed to discredited testimony may not simply return to baseline, but instead, may overcorrect in their judgments 'bending over backwards' to be fair" (p. 492).

In sum, some of the findings described above lend support to the idea that although there is a general tendency to accept eyewitness testimony uncritically, discrediting an eyewitness in court will tend to undermine the influence of that witness on the decision-making process.

5.3 Witness confidence and accuracy

One strategy for discrediting a witness is to confuse the witness such that the witness becomes less confident in his or her reports. This is significant in that the confidence shown by witnesses is noted to be one of the main factors driving jurors' perceptions of witness credibility and accuracy (Berman, Narby & Cutler, 1995; Brewer, Potter, Fisher, Bond & Luszcz, 1999; Wells, 1985; Wells, Lindsay & Ferguson, 1979). As mentioned earlier, there is a variety of evidence to indicate that jurors intuitively believe that confidence is a valid predictor of accuracy; hence many studies have demonstrated that jurors and jurists rely heavily upon the demeanour of the witness when judging accuracy;

if the witness appears to be confident, he or she will be considered more accurate (Brigham & Wolfskeil, 1983; Cutler & Penrod, 1988; Cutler & Penrod, 1995; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Stuve, 1988; Deffenbacher & Loftus, 1982; Fox & Walters, 1986; Kassin, Rigby and Castillo, 1991; Leippe, Manion, & Romanczyk, 1992; Lindsay, 1994; Lindsay, Glenn, Nosworthy & Martynuck, 1994; Lindsay, Wells & O'Connor, 1989; Lindsay, Wells & Rumpel, 1981; Sporer, 1992; Luus & Wells, 1994; Wells, Ferguson & Lindsay, 1981; Wells, Lindsay & Ferguson, 1979; Whitely & Greenberg, 1986). In contrast when, in 1989, Kassin, Ellsworth and Smith surveyed 63 experts on eyewitness testimony 87% stated that they would testify that 'eyewitness confidence is not a good predictor of his or her accuracy'. More recent research, however, suggests that the issue is somewhat more complex.

Reviews assessing witness C-A relationships have suggested that C-A varies significantly from one study to another. Although a few negative relationships have been reported, the general tendency is for there to be either no relationship, or a small positive relationship between eyewitnesses' confidence and their accuracy (Bothwell, Deffenbacher, & Brigham, 1987; Deffenbacher, 1980; Fruzzetti, Toland, Teller & Loftus, 1992; Perfect, Watson & Wagstaff, 1993; Nelson, 1988; Penrod, Loftus & Winkler, 1982; Penrod, Read, & Cutler, 1995; Wells, 1993; Wells & Murray, 1984). However, in addition, Sporer, Penrod, Read and Cutler (1995) found that including choice as a moderator variable leads to a different conclusion: for those making a positive identification (choosers), the C-A correlation was reliably and consistently higher than for non-choosers.

Deffenbacher (1980) proposed that variability in C-A relationships could be explained, to some degree, by the 'optimality' of the conditions present at the time information processing occurs; at encoding, storage, and retrieval. Deffenbacher claims that, as fewer errors in judgement are elicited by more optimal conditions, these coincide with strong positive C-A relationships. Zero or negative C-A relationships, however, are more likely to exist under conditions that are non-optimal, and which are more likely to produce mistakes.

Leippe (1980) also stressed the importance of memory, cognitive and social factors that may influence participants' confidence. Leippe argues that witnesses will display negative, low or nonexistent C-A relationships when they are subject to processes that may alter memory and cognitions without their awareness. Moreover, this is most likely to happen when conditions are non-optimal. Arguably the subtle, less transparent, questioning strategies used in cross-examination, might act in precisely this way leading to lower, zero, or even negative relationships between confidence and accuracy.

5.3.1 Methodology

Another approach to identifying the source of differences in C-A relationships has been to examine the methods used to assess memory and calculate C-A. For example, Robinson and Johnson (1996) reported that free recall memory conditions resulted in higher eyewitness C-A correlations than recognition memory conditions. In another study, Smith, Kassin and Ellsworth (1989), calculated within-subjects' C-A relationships (i.e. a separate C-A relationship for each participant; clearly a witness may be clear and confident about a

particular issue but not about others) and between-subjects' C-A relationships (i.e. calculating a single C-A relationship over a number of witnesses). They used a slide presentation followed by two-alternative forced choice questions. Participants were then required to rate their confidence for each response indicated on a ten-point scale. Smith *et al.* reported low (r = <.2) correlations between confidence and performance for ratings taken both within and between subjects. They conclude:

> "Confidence is not a good predictor of accuracy. Commonsense 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 witnesses' 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) criticised Smith *et al.'s* use of a two-choice recognition procedure in assessing memory, which resulted in a hit rate of 63%. They argued that as 37% of the guesses were incorrect by chance, the implication is that 37% of the hits would have been correct by chance. To reduce the chances of eliciting a hit through guessing Perfect *et al.* used a five-alternative forced choice question procedure. A higher overall correlation was reported between-subjects (Goodman-Kruskal Gamma = .49) than Smith *et al.*, but no correlation (Gamma = -.03) was

reported for the within-subjects analysis. Perfect *et al.* (1993) also mentioned the difficulty of items as problematic for Smith *et al.* in drawing their conclusions regarding C-A relationships.

5.3.2 Item difficulty

One possible reason for low C-A relationships in laboratory studies is that researchers tend to select items to avoid ceiling and floor effects (i.e. avoiding items that are either very easy or very hard to remember). In effect, this may reduce the variance necessary to produce high and significant C-A correlations. For example, in terms of questions relating to identification, gender has been shown likely to be answered accurately, whereas eye colour is less likely to be answered accurately (Christianson & Hubinette, 1993); however, in many studies, because of these very characteristics such items might be avoided. In view of this rationale, Kebbell, Wagstaff and Covey (1996) investigated the influence of item difficulty on the C-A relationship. Kebbell *et al.* reported that when questions were used that included items varying in difficulty, thus maximising the probabilities of producing 'absolutely certain' and 'pure guess' responses, C-A relations (both between and within subjects) were generally considerably higher than those previously reported by Smith *et al.* (1989) and Perfect *et al.* (1993). Furthermore, when participants were 'absolutely certain' that the information they gave was correct, they invariably were accurate. These findings appear to support the idea that previous research had chosen unrealistic 'moderate' items, reducing the variance necessary to achieve high correlations.

5.4 Conclusion

To summarize, it seems that, in general both jurists and jurors tend to be rather over-credulous in the accuracy of eyewitness' testimony, though discreditation through cross-examination may be effective. A major factor in the determination of eyewitness accuracy is the confidence that witnesses appear to display, both in demeanour and in their reports. Hence, there is also a tendency for people to believe that confident witnesses are most accurate. Evidence is mixed however, with regard to the actual relationship between eyewitness confidence and accuracy. On the whole the relationship seems to be positive but low, however, quite significant relationships can be found when conditions are optimal and items are easy to remember, or items vary considerably in difficulty such that both easy and difficult items are included.

Given the potential impact of C-A relationships on the judicial process it would seem very important to assess the effects of cross-examination styles on these relationships; for example, it is possible that a confusing lawyerese style might disrupt C-A relationships by making witnesses less sure about their testimony than they might otherwise have been. This, in turn might be moderated by factors such as question difficulty (most disruption would be found for difficult items), and method of calculation (most disruption would be found for within subjects C-A correlations).

CHAPTER 6

INTRODUCTION TO THE EXPERIMENTAL WORK

Given that the central focus of this thesis is on the effects of courtroom questioning styles, numerous issues arise that are worthy of systematic empirical investigation; nevertheless, on the basis of the literature reviewed so far, a number of key factors emerge as useful starting points for a research programme in this area. These are as follows.

6.1 Cross-examination questioning styles

Chapter 2 showed that linguistic elements contained within sentence structures can influence verbal reports. Thus, questioning styles inherent to cross-examination might influence witnesses' verbal evidence in terms of the accuracy and confidence of reports given, and in respect of the relationship between confidence and accuracy (C-A). In particular, 'lawyerese' styles that employ assertive and leading questions may lead to reductions in the number of accurate responses made by witnesses, and the confidence expressed in the answers made.

Additionally, when negative feedback is used during cross-examination the inference is that the witness's first answer is incorrect, thus such feedback, either applied explicitly or implicitly, implies an alternative response is more appropriate. The expectation upon witnesses to produce different answers might lead them to become generally uncertain about the accuracy of their responses.

Furthermore, as the cross-examination interview includes items that cover a range of difficulty, and cross-examiners tend to target those items that are perhaps the least memorable, it seems appropriate to investigate not only the influence of cross-examination questioning styles and strategies on witness accuracy, confidence, and the relationship between confidence and accuracy, but also to examine how item difficulty might affect these variables. In particular, the evidence reviewed in Chapter 5 suggests that one might expect any negative effects on accuracy and confidence to be most apparent with difficult items.

6.2 Delay and the stability of confidence and accuracy

Within the Criminal Justice System, the time delay between observations initially made by witnesses (i.e. by police statement) and the presentation of evidence in court can at best be months, at worst years. The same applies to giving testimony in court on more than one occasion; as in cases of appeal. Given that memory inevitably declines over time, one would expect that reductions in accuracy and confidence might be observed in such situations. However, there is, as yet, no evidence in the literature concerning the effects of time on performance under cross-examination. One possibility is that the relationship between confidence and accuracy might remain stable through the consolidation of memory processes. For example, where a witness is initially uncertain about particular items, over time, memory traces for such items present might decay leaving no memories for these items. Conversely, where a witness is very confident about particular items, these traces remain stable or increase in strength over time through rehearsal, such that these items are

clearly memorable later. The effect of any such stabilisation process would be to maintain the relationship between confidence and accuracy, even although overall levels of accuracy and confidence might decline.

6.3 Jurors' perceptions of witnesses

In Chapter 5 it was also noted that both judges and jurists are more likely to believe that witnesses are telling the truth if witnesses express clear confidence in answers given. As a result, different examination styles might lead jurors to make different inferences with regard to witness credibility. For instance, the inclusion of negative feedback might serve to lower the perceived confidence of the witness and make observers doubt the accuracy of the witness' testimony.

6.4 Predicting a good witness under cross-examination

The literature reviewed in Chapter 3 indicated that individual differences, in particular those relating to suggestibility, might be important in determining witnesses' responses to cross-examination interview styles and strategies. This evidence in turn invites the question, who might make a good or bad witness under cross-examination? One obvious approach to this issue would be to administer the Gudjonsson Suggestibility Scales (GSS) to all witnesses in the present studies. However, this option was not used in the present thesis for a number of reasons. Most important, there is a large database on the GSS and the use of the GSS would add little to our knowledge in this area. The GSS essentially works on the basis of negative feedback, one of the main conditions in the research presented here. It would not be surprising,

therefore, if performance on both tasks was related. Second, if given before the experimental conditions, the GSS might prime participants as to the nature of the variables assessed, and, if given afterwards, the experimental conditions might prime participants as to the responses expected on the GSS. And, third, if the present research is to be used to develop a quick and reliable possible tool for screening potentially good and poor witnesses in court, the GSS is too time consuming, requires trained clinical judgment to interpret scores, yet from the witnesses point of view, is very transparent in terms of its purpose. Nevertheless, the literature suggests that there are a number of variables that might usefully be assessed as possible predictors of individual differences in eyewitness performance; for example, cognitive capacity and the experience of stress.

6.5 Aims of the empirical research

The considerations outlined above give rise to four questions addressed in the current thesis.

1) Do cross-examination questioning styles, including the provision of negative feedback, affect witness confidence and accuracy, and the relationship between confidence and accuracy (C-A)?

2) Do accuracy, confidence, and the C-A relationship, so influenced, remain stable over time?

3) Do jurors assess the accuracy and confidence of witnesses differentially according to the way in which they are examined?

4) Can we predict who will be a good and poor witness under cross-examination?

PART 2

THE EMPIRICAL RESEARCH

CHAPTER 7

CROSS-EXAMINATION QUESTIONING STYLES: Experiment 1

7.1 INTRODUCTION (overview)

As mentioned at the beginning of this thesis, the purpose of cross-examination is to challenge witnesses' evidence and their credibility, and opposing counsel are frequently allowed to use specific tools such as leading and highly suggestive forms of questioning (lawyerese questions). However, although some legal advisors openly assert that asking leading questions can be unwise, others contend that they form a normal, useful, and effective procedure for verifying doubtful information and producing new information (Evans, 1995). In contrast, a number of studies have shown that such question types typically employed during cross-examination may actually impede witness accuracy (Kebbell & Giles, 2000), and unwittingly provide obstructions to the truth (Westcott, 1995).

However, as shown in Chapter 2 the evidence on this issue is mixed. For example, Dodd and Bradshaw (1980) found that such effects could be cancelled if the person being questioned perceived the questioner to be biased, and was thus suspicious of the intentions of the questioner; i.e. if witnesses adopt a 'suspicious set' (Gudjonsson, 1989; 1992). Similar effects have been demonstrated in respect of witness confidence. Wells, Ferguson and Lindsay (1981) found that witnesses who had been 'briefed' through rehearsal of their

testimony that a cross-examiner will look for inconsistencies were likely to show inflated confidence.

Another questioning strategy sometimes employed in cross-examination is that of negative feedback; that is, the witness may be asked questions that imply that his or her first answer may be incorrect; for example, "think about this again"; "try and be more accurate"; "consider your answer carefully", and so on, though little has been said of this in the legal literature. According to Schooler and Loftus (1986), through the principle of *discrepancy detection* negative feedback may reduce witnesses' confidence in their own memories and make them more suggestible to leading questions. However, as yet, no research has been published on the effects of negative feedback as might be given in a courtroom situation.

The possibility that questioning styles may affect confidence is particularly important given that confidence is one of the main factors, if not the main factor, driving jurors' perceptions of witness credibility (Lindsay, 1994). However, evidence on this topic is also mixed. Whilst some studies have reported fairly robust positive relationships between confidence and accuracy (Kebbell, Wagstaff and Covey, 1996; Perfect, Watson & Wagstaff, 1993), others have not (Fruzzetti, Toland, Teller & Loftus, 1992; Wells, 1993). The aim of the first study, therefore, was to examine the effects of cross-examination questioning styles on witness accuracy and confidence.

7.1.1 Hypotheses

More specifically, it was predicted that; a) leading questions would reduce accuracy, confidence, and the relationship between accuracy and confidence; b) these negative effects would be exacerbated by asking leading questions in a declarative form (i.e. one that strongly and openly implies that the facts upon which the question is based are correct or incorrect); and c) progressive reductions in accuracy, confidence, and the relationship between confidence and accuracy would occur as questions were asked in a confusing 'lawyerese' form, and, d) in a lawyerese form accompanied by negative feedback.

7.2 METHOD

7.2.1 Participants

The participants were 24 males and 36 females who were either students at the University of Liverpool or prospective students. Mean age was 23.08 (range = 17-50). Participants were randomly assigned to one of three conditions: 'Control' (N=20), 'Lawyerese' (N=20) or 'Negative Feedback' (N=20).

7.2.2 Materials and procedure

Participants were then told the study concerned eyewitness memory; they were to observe a film for 4-5 minutes and afterwards had to answer some questions. They were then shown a colour video of an event during which a woman at a bus stop is abducted and forced into a car; a gun was used.

After this, participants were required to complete a filler task for a period of 5 minutes, which involved reading unrelated material. Participants

were advised that they would now be asked a series of questions regarding the videotape. The questions were randomised within the interview schedule and these allocations remained stable throughout the conditions. All questions required 'yes' or 'no' answers, and following each question participants were asked to rate their confidence in the response they had given. A Likert scale ranging from (1) 'pure guess' to (9) 'absolutely certain' was used. As a precautionary measure, the interviewer, who was also the experimenter, was trained to give similar inflexions. This was achieved by recording several interviews conducted by the interviewer, who then underwent trials to learn to modulate the voice consistently throughout the interviews. The three conditions were as follows:

In the *Control condition* participants were required to give answers to 42 simple questions, in three sets of 14; (a), (b), and (c). For example, 'did the younger man have a moustache?' Within each set of questions, half of the questions required a 'yes' response and half required a 'no' response to be accurate (see Appendix 1).

In the *Lawyerese condition* three kinds of questions were asked. 14 were simple, 14 leading and 14 declarative. The 14 simple questions were identical to the 14 items in set (a) of the simplified condition. The leading questions were so framed as to prompt the answer desired, for example, 'a bus arrived up the road shortly before the attack, didn't it?' The declaratives were highly leading questions requiring an affirmative response, for example; 'you've seen the gang consisted of two men'. The leading and declarative questions were derived from sets (b) and (c) in the control condition,

respectively, for example, 'did a bus arrive up the road shortly before the attack?' and 'did the gang consist of two men?' (see Appendix 2).

In the *Negative Feedback condition* the questions were the same as those used in the lawyerese condition (14 simple, 14 leading and 14 declarative), however, negative feedback was applied to each leading and declarative question in order to prompt another 'yes' or 'no' response (for example; 'Try harder to remember; the younger man, did he have glasses on?'). Thus, participants were told that some might receive further questions about the answers they had given. Confidence ratings were required after the feedback (see Appendix 3).

At the end of the interview schedule, participants were thanked for their participation and debriefed. During this time, an explanation of the rationale for the experiment was outlined and participants were given the opportunity to ask questions and raise issues for discussion. Many of the participants talked about relevant issues which both concerned and interested them. These are highlighted in the discussion (see p. 189).

7.3 RESULTS

Accuracy in terms of the number of items correct was compared for each condition (i.e. Control, Lawyerese, and Negative Feedback) separately, by three one-way ANOVAs (i.e. one for question set a, one for question set b, and one for question set c). No significant effects were found in any of the analyses ($p>.1$). The means and standard deviations are shown in Table 7.3.1. Similar analyses on the confidence ratings also showed no significant

differences between the three conditions in any of the three ANOVAs ($p>.1$), see Table 7.3.2.

Table 7.3.1 Mean accuracy with respect to questioning style and question type set

STYLE	QUESTION TYPE SET Accuracy		
	Set a	**Set b**	**Set c**
Control N=20	**10.20** (Simple) (1.64)	**8.75** (Simple) (1.83)	**9.90** (Simple) (1.77)
Lawyerese N=20	**9.35** (Simple) (2.01)	**8.75** (Lead'g) (1.25)	**10.00** (Decl've) (1.69)
Negative Feedback N=20	**9.35** (Simple) (1.69)	**8.95** (Lead'g) (1.43)	**10.45** (Decl've) (1.32)

Note: Standard deviations are shown in brackets

Table 7.3.2 Mean confidence with respect to questioning style and question type set

STYLE	QUESTION TYPE SET Confidence		
	Set a	**Set b**	**Set c**
Control N=20	**5.57** (Simple) (1.12)	**5.67** (Simple) (.92)	**5.79** (Simple) (1.00)
Lawyerese N=20	**5.56** (Simple) (.84)	**5.59** (Lead'g) (.73)	**6.04** (Decl've) (1.03)
Negative Feedback N=20	**5.41** (Simple) (.78)	**5.41** (Lead'g) (.76)	**5.49** (Decl've) (1.19)

Note: Standard deviations are shown in brackets

Analyses were then performed on the 21 questions in each condition that contained incorrect information and required a 'no' answer to be correct. Again, accuracy in terms of the number of items correct was compared for each condition (i.e. Control, Lawyerese, and Negative Feedback) separately, by three one-way ANOVAs (i.e. one for question set a, one for set b, and one for set c). No significant differences were found in any of the analyses ($p>.1$). The means and standard deviations are shown in Table 7.3.3. However, trends were observed for both set b, containing leading questions, $F(2,57) = 2.44$, $p = .096$, and set c, containing declarative questions $F(2,57) = 2.86$, $p = .065$. In each case the most accurate answers were given in the Negative Feedback condition, and least accurate in the Control condition. Similar analyses on the confidence ratings also showed no significant differences between the three

conditions in any of the three ANOVAs (p>.1); see Table 7.3.4.

Table 7.3.3 Mean accuracy for questions requiring a 'no' answer with respect to questioning style and question type set

STYLE	QUESTION TYPE SET Accuracy		
	Set a	**Set b**	**Set c**
Control N=20	**4.65** (Simple) (1.23)	**3.80** (Simple) (1.67)	**4.40** (Simple) (1.39)
Lawyerese N=20	**4.55** (Simple) (1.43)	**4.35** (Lead'g) (1.04)	**4.85** (Decl've) (1.23)
Negative Feedback N=20	**4.80** (Simple) (1.20)	**4.70** (Lead'g) (1.08)	**5.35** (Decl've) (1.14)

Note: Standard deviations are shown in brackets

Table 7.3.4 Mean confidence for questions requiring a 'no' answer with respect to questioning style and question type set

STYLE	QUESTION TYPE SET Confidence		
	Set a	**Set b**	**Set c**
Control N=20	**4.74** (Simple) (1.43)	**5.90** (Simple) (1.06)	**5.39** (Simple) (1.23)
Lawyerese N=20	**4.76** (Simple) (1.36)	**5.67** (Lead'g) (.83)	**5.51** (Decl've) (1.52)
Negative Feedback N=20	**4.56** (Simple) (1.34)	**5.76** (Lead'g) (.99)	**5.50** (Decl've) (1.29)

Note: Standard deviations are shown in brackets

Within subjects confidence-accuracy (C-A) correlations were calculated using Pearsons (all correlations are Pearsons unless otherwise specified) for each participant across the 42 questions, producing three correlations for each participant, one for question set a, one for question set b, and one for question set c, based on 14 questions each. These correlations were then compared for each condition (Control; Lawyerese; Negative Feedback) separately, by three one-way ANOVA's (i.e. one for question set a, one for set b, and one for set c). No significant differences were found in any of the analyses ($p>.1$); see Table 7.3.5.

Table 7.3.5 **Mean 'within-subjects' C-A correlations (including overall means) with respect to questioning style and question type set**

STYLE	QUESTION TYPE SET		
	Set a .12	**Set b** .22	**Set c** .14
Control .12 N=20	**.10** (Simple) (.27)	**.16** (Simple) (.34)	**.10** (Simple) (.23)
Lawyerese .14 N=20	**.05** (Simple) (.22)	**.27** (Lead'g) (.20)	**.11** (Decl've) (.33)
Negative Feedback .22 N=20	**.21** (Simple) (.26)	**.24** (Lead'g) (.21)	**.22** (Decl've) (.27)

Note: Standard deviations are shown in brackets

The C-A correlation was then calculated for each cell in table 7.3.5, using the mean confidence and correct scores for each participant as data (between-subjects). These between-subjects C-A relationships were then tested for difference following transformation to produce a *z* statistic. No significant differences were observed; see Table 7.3.6.

Table 7.3.6 'Between-subjects' C-A correlations with respect to questioning style and question type set

STYLE	QUESTION TYPE SET		
	Set a	**Set b**	**Set c**
Control .06 N=20	**.22** (Simple)	**.02** (Simple)	**.10** (Simple)
Lawyerese .23 N=20	**.09** (Simple)	**.25** (Lead'g)	**.14** (Decl've)
Negative Feedback -.03 N=20	**-.19** (Simple)	**.07** (Lead'g)	**.26** (Decl've)

Note: Standard deviations are shown in brackets

Within-subjects confidence-accuracy (C-A) correlations were also calculated for each participant for the 21 questions requiring a 'no' answer (7 from set a, 7 from set b, and 7 from set c). These correlations were then compared for each condition (Control; Lawyerese; Negative Feedback) separately, by three one-way ANOVA's (i.e. one for question set a, one for set b, and one for set c). No significant differences were found in any of the analyses ($p>.1$); see Table 7.3.7.

Table 7.3.7 Mean 'within-subjects' C-A correlations for questions requiring a 'no' answer (including overall means) with respect to questioning style and question type set

STYLE	QUESTION TYPE SET		
	Set a -.12	**Set b** .03	**Set c** -.08
Control -.13 N=20	**-.16** (Simple) (.31)	**-.04** (Simple) (.48)	**-.19** (Simple) (.35)
Lawyerese -.05 N=20	**-.19** (Simple) (.30)	**.08** (Lead'g) (.29)	**-.05** (Decl've) (.46)
Negative Feedback .00 N=20	**-.03** (Simple) (.39)	**.04** (Lead'g) (.34)	**.01** (Decl've) (.29)

Note: Standard deviations are shown in brackets

The mean confidence in 'incorrect' answers was then compared with mean confidence in 'correct' answers for each condition (i.e. Control, Lawyerese, and Negative feedback) for each question set, separately, by means of three two-way mixed ANOVAs (i.e. one for question set a, one for set b, and one for set c), with repeated measures on the second factor (Control; Lawyerese; Negateve feedback X Confidence in Incorrect; Confidence in Correct). In each case no main effects between groups were found ($p>.1$). However, significant main effects for confidence was observed; in each case, confidence in correct answers was significantly higher than confidence in incorrect answers. The F(1,57) values for set a, set b, and set c, were; 13.24, $p = .001$; 46.64, $p = .000$; and 14.79, $p = .000$. The means and standard deviations are

shown in Tables 7.3.8, 7.3.9, and 7.3.10. No other significant effects were observed.

Table 7.3.8 Mean confidence in correct and incorrect answers (including overall means) to question set a by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
5.40 (1.09)	**5.73** (.94)	**5.07** (1.58)
Control		
N=20 **5.51** (1.33)	**5.81** (.98)	**5.21** (1.94)
Lawyerese		
N=20 **5.50** (.91)	**5.71** (.85)	**5.30** (1.27)
Negative Feedback		
N=20 **5.18** (1.02)	**5.66** (1.03)	**4.70** (1.45)

Note: Standard deviations are shown in brackets

Table 7.3.9 Mean confidence in correct and incorrect answers (including overall means) to question set b by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
5.42 (.84)	**6.06** (.92)	**4.78** (1.26)
Control		
N=20 **5.53** (.98)	**6.06** (.97)	**4.99** (1.59)
Lawyerese		
N=20 **5.39** (.77)	**6.13** (.85)	**4.64** (1.22)
Negative Feedback		
N=20 **5.34** (.77)	**5.98** (1.00)	**4.71** (.91)

Note: Standard deviations are shown in brackets

Table 7.3.10 Mean confidence in correct and incorrect answers (including overall means) to question set c by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
5.66 (1.20)	**6.10** (1.08)	**5.21** (1.81)
Control		
N=20 **5.71** (1.10)	**6.01** (.99)	**5.41** (1.59)
Lawyerese		
N=20 **5.84** (1.27)	**6.29** (.97)	**5.37** (2.01)
Negative Feedback		
N=20 **5.42** (1.24)	**5.99** (1.28)	**4.85** (1.84)

Note: Standard deviations are shown in brackets

Finally, out of 840 possible responses given by the Control group, 197 were rated in terms of confidence as 'absolutely certain', of these, 76% were correct. For the Lawyerese group, 210 were rated as 'absolutely certain', and of these, 80% were correct, and for the Negative Feedback group, 175 were rated as 'absolutely certain', of which 86% were correct. Overall, 80% of 'absolutely certain' answers were correct.

7.4 DISCUSSION

There were no significant effects for accuracy or confidence between the three questioning style conditions (Control, Lawyerese and Negative Feedback). Indeed, what trends there were seemed to be contrary to predictions; i.e. the tendency was for most accuracy to be shown in the Negative Feedback condition and least in the Control condition. With regard to accuracy these results appear to conflict with those that suggest question styles typically employed during cross-examination may actually impede witness accuracy (Kebbell & Giles, 2000; Westcott, 1995).

In terms of C-A correlations also, the means did not suggest that C-A relationships were reduced in the Lawyerese and Negative Feedback conditions; indeed, higher C-A relationships were shown in the Negative Feedback condition. Even the highest number of absolutely confident/correct answers was shown in the Negative Feedback group. According to Schooler and Loftus (1986), negative feedback reduces witnesses' confidence in their memories and thus they are less likely to compare the suggestions made by a questioner with information in memory. This clearly did not occur in the present study. One possible reason for this is that the use of negative feedback might actually have helped witnesses to check memories for an event with information contained within the question.

In keeping with much previous research, as regards C-A in general, C-A relationships tended to be low, but there were significant differences in confidence between correct and incorrect answers for all question type sets; witnesses were more confident in correct answers. Also when participants were absolutely certain of their response, they tended to be correct (80%

overall; see also, Kebbell, Wagstaff & Covey, 1996). Predictably, C-A correlations for questions requiring an incorrect response) were particularly low; indeed they tended to be negative.

To summarize, the results of the present study suggest that the scepticism that some linguists and legal commentators have displayed regarding the questioning styles and type of questions used in courtroom situations may to some extent have been misplaced. Of course, the experimental set up used here cannot reflect the full range of variables present in an actual cross-examination, nevertheless, it could be argued that the use of individual interviewing in the present study more accurately reflects what happens in a courtroom, than the anonymous group questionnaires used in some previous studies. On face value, therefore, the present results suggest that one cannot conclude that even highly leading questions will, as a matter of course, have a detrimental effect on eyewitness confidence and accuracy in an interrogative situation.

CHAPTER 8

CROSS-EXAMINATION QUESTIONING STYLES: Experiment 2

8.1 INTRODUCTION

Results from the first study suggest that questioning styles used in cross-examination may not have the negative impact shown by some previous studies. However, there may be a number of reasons why no significant effects were found.

For example, the more overt kind of feedback used in the first study may have appeared transparent in its function and created suspicion in witnesses, possibly negating any effects on accuracy and confidence (Gudjonsson, 1992). However, not all negative feedback used during actual cross-examination is as explicit. A more covert and subtle approach might consist of re-directing questions such as, "is it possible that you might be mistaken, and that ....?", and so on; though, as mentioned previously, little has been said of feedback strategies *per se* in the legal literature.

Also, the random allocation of question items to questioning style conditions, together with the requirement for equal numbers of 'yes' and 'no' answers in each questioning condition, may have created further suspicion. For example, randomly allocating an 'easy' item to a declarative question form (i.e. "you've seen it was dark when the attack occurred?") requiring a 'yes' response that is clearly inaccurate (i.e. it was actually daylight), would almost certainly have signalled the purpose of the interview to the witness.

In sum, despite some attempts to accurately reflect what happens in a

courtroom (e.g. through the use of individual interviewing), the strict experimental method used in the first study may still have constrained its ecological validity. In view of these considerations, a second study was conducted to examine effects of cross-examination questioning styles on eyewitness accuracy and confidence, but this time the interview was specifically designed and constructed so as to replicate as far as possible the way in which counsellors might conduct cross-examinations.

In addition, item difficulty was considered. As noted previously, according to Kebbell, Wagstaff & Covey (1996), one of the crucial variables affecting witness accuracy and confidence may be item difficulty; thus little correlation will be found between C-A when witnesses are asked to remember a relatively homogeneous pool of difficult items (for example, items concerning peripheral detail; artefacts such as mailboxes, pictures); however, if variance is increased through the inclusion of items that are easier to remember (sex, whether it was day or night), confidence-accuracy relationships will improve. Although the items in the first study were selected to be relatively heterogeneous, in the following study an attempt was made to investigate item difficulty more systematically.

8.1.1 Hypotheses

It was predicted that, a) questions asked in a standard lawyerese style would reduce accuracy, confidence, and the relationship between confidence and accuracy; b) further reductions in accuracy and confidence would occur when the lawyerese questioning style was accompanied by subtle negative feedback, and c) these effects would be most evident for difficult items.

8.2 METHOD

8.2.1 Participants

The participants were 12 males and 48 females drawn from an opportunity sample. All participants were either students at the University of Liverpool or members of a research panel in the Department of Psychology at the University. The population mean age was 23.13 (SD = 5.96; range = 18-52).

8.2.2 Materials and procedure

Participants were randomly assigned to one of three conditions; 'Control' (N=20), 'Lawyerese' (N=20) or 'Lawyerese with Negative Feedback' (N=20). All participants were then told that they were to observe a videotape of an event sequence for a period of around 4-5 minutes, and afterwards they would be required to answer some questions. All participants were then shown the five minutes colour video depicting a criminal offence used in experiment 1.

After this, every participant was required to complete a filler task for a period of five minutes, which involved reading unrelated material. Participants were advised that they would then be asked a series of questions regarding the videotape they had observed. All questions required 'yes' or 'no' answers, and following each question participants were asked to rate their confidence in the response they had given on a Likert scale ranging from 1 to 9, where (1) represented 'pure guess' and (9) represented 'absolutely certain'. The interviewer was again trained to give similar inflexions throughout the interviews using the same method as detailed in experiment 1.

Each participant was then individually interviewed under one of the

following three conditions.

In the *Control condition* participants were required to give answers to 38 simple questions, in three sets varying in item difficulty (i.e. ease with which they can be remembered): (a) represented 14 'easy' target items, (b) represented 14 'moderately difficult' target items, and (c) consisted of 10 'difficult' items. Examples were, 'did two men carry out the attack on the victim?'; 'did the victim have long hair?'; and 'would you say that this car had four doors?', for the easy, moderate, and difficult categories, respectively (see Appendix 4). Item difficulty was determined using the accuracy data from experiment 1.

In the *Lawyerese condition* participants were again required to give answers to the same three sets of target items as in the Control condition (i.e. 14 'easy', 14 'moderate' and 10 'difficult' items), but the questions were so phrased as to replicate the manner in which lawyers conduct cross-examinations in court. The phrasings were taken directly from the examination of several Crown Court transcripts. For example, with regard to the items previously mentioned, the questions became, 'do you also remember that two men carried out the attack on the victim?'; 'you would agree that the victim's hair was long?'; and 'isn't it also right this car had four doors?', for the easy, moderate, and difficult categories, respectively (see Appendix 5). These phrases prompted the answer desired and required more affirmative responses than in the Control condition.

In the *Lawyerese with Negative Feedback condition* exactly the same schedule was employed as in the Lawyerese condition (i.e. 14 'easy', 14 'moderate' and 10 'difficult' items), however, subtle negative feedback was

applied to each 'no' answer, regardless of accuracy, in order to reassert the expectancy of 'yes' responses; this was done by reasserting the question preceded by the statement 'is it a possibility that you might be mistaken and that...etc.?'; see Appendix 6).

To maintain ecological validity the schedule was specifically constructed with an imbalance between the numbers of items in the item difficulty categories, and the requirement for 'yes' and 'no' responses. A detailed examination and analysis of the transcripts indicated that a minimum of 82% of the questions asked during cross-examination required a 'yes' or 'no' answer, and that 71% of questions asked could be classified as closed and leading or heavily leading. Significantly, the transcripts clearly showed that, in practice, lawyers tend to phrase the majority of questions such that a 'yes' response is encouraged, thus making it more difficult for witnesses to switch to a 'no' response on the fewer critical items. In general, critical items are those that are associated with inconsistency both within and between witnesses' statements. As a consequence, critical items are items about which there may be some ambiguity, and this is most likely to be the case for items that are difficult to remember. From the cross-examining lawyers' point of view, the strategy of targeting difficult items makes sense, as cross-examiners would be considerably less likely to influence the witness' response to a very easy item. Also, it is also clear that lawyers will tend to target only a few critical items to prevent their strategy from becoming too transparent. Consequently, when wishing to cast doubt on the reliability of a particular witnesses' testimony overall, they will target a few difficult items and set these up as exemplars. This procedure was therefore adopted in the present

study. Consequently, in all conditions for the 'easy' and 'moderate' items a 'yes' response was required to be accurate (i.e. 28 'yes' responses were required), while the 'difficult' targets required a 'no' answer to be correct (i.e. 10 'no' responses were required).

At the end of the interview schedule, all participants completed a series of questionnaires (see Appendix 7), were thanked for their participation and debriefed. At this time it was appropriate to give an explanation of the rationale for the experiment to each participant and also create the opportunity for participants to ask questions, which in turn raised issues for discussion (highlighted again in the discussion on p. 189).

The nature and results of the questionnaires are reported in Chapter 12.

8.3 RESULTS

One-way ANOVAs (Control, Lawyerese, Lawyerese with Negative Feedback) were performed using the overall data (easy, moderate and difficult items combined) for accuracy (mean number of items correct), confidence (mean scores on the 1-9 Likert scale), and within subjects confidence-accuracy correlations (C-A; mean of the confidence accuracy correlations for each participant), calculated using Pearsons (all correlations are Pearsons unless otherwise specified). None of the main effects was significant for any of these analyses. The only significant effect to emerge from the combined data was a significant between subjects confidence-accuracy correlation (i.e. the correlation between the total accuracy score and the total confidence score for each witness) for the Lawyerese with Negative Feedback condition (r=.46,

n=20, $p < .05$); the same correlations for the Control and Lawyerese conditions were positive but not significant, though the correlations were not significantly different from each other (z statistic, $p > .05$). The means and standard deviations are shown in Table 8.3.1.

Table 8.3.1 Mean accuracy, confidence, C-A (w-s), and C-A (b-s) correlations for 'All' questions by style of questioning

STYLE	Accuracy	Confidence	C-A (w-s)	C-A (b-s)
Control N=20	**26.85** (1.95)	**5.93** (.85)	**.39** (.13)	**.18**
Lawyerese N=20	**26.05** (2.76)	**5.86** (.75)	**.35** (.12)	**.38**
Lawyerese with Negative Feedback N=20	**25.65** (2.72)	**5.59** (.68)	**.42** (.13)	**.46***

Note: Standard deviations are shown in brackets; *$p < .05$

One-way ANOVAs (Control, Lawyerese, Lawyerese with Negative Feedback) were then performed separately on the 14 'easy', the 14 'moderate', and the 10 'difficult' questions for each condition. Again, separate analyses were conducted for accuracy, confidence, and confidence-accuracy correlations (mean of the confidence-accuracy correlations for each participant).

No significant main effects for the three styles of questioning were found for the 'easy' items; see Table 8.3.2. Again, the only significant effect

to emerge from the data for 'easy' items was a significant between subjects confidence-accuracy correlation (i.e. the correlation between the total accuracy score and the total confidence score for each witness) for the Lawyerese with Negative Feedback condition (r =.71, n = 20, p< .001); the same correlations for the Control and Lawyerese conditions were again positive and non-significant, and none differed significantly from each other (z statistic, p>.05). The means and standard deviations are shown in Table 8.3.2.

Table 8.3.2 Mean accuracy, confidence, C-A (w-s), and C-A (b-s) correlations for 'Easy' questions by style of questioning

STYLE	Accuracy	Confidence	C-A (w-s)	C-A (b-s)
Control N=20	**11.75** (1.45)	**7.26** (.73)	**.44** (.29)	**.36**
Lawyerese N=20	**12.10** (1.02)	**6.97** (.81)	**.54** (.26)	**.33**
Lawyerese with Negative Feedback N=20	**12.50** (1.24)	**7.02** (.81)	**.45** (.29)	**.71*** **

Note: Standard deviations are shown in brackets; ***p<.001

No significant main effects for the three styles of questioning were found for the 'moderate' items either; see Table 8.3.3. However, a significant between subjects confidence-accuracy correlation (i.e. the correlation between the total accuracy score and the total confidence score for each witness) was found

again but this time for 'moderate' items in the Lawyerese condition (r=.61, n=20, p< .01); the same correlations for the Control and Lawyerese with Negative Feedback conditions were positive but not significant, and none of the correlations differed significantly from each other (z statistic, p>.05). The means and standard deviations are illustrated in Table 8.3.3.

Table 8.3.3 Mean accuracy, confidence, C-A (w-s), and C-A (b-s) correlations for 'Moderate' questions by questioning style

STYLE	Accuracy	Confidence	C-A (w-s)	C-A (b-s)
Control N=20	**10.05** (1.36)	**5.26** (1.01)	**.46** (.28)	**.04**
Lawyerese N=20	**9.10** (2.25)	**5.39** (1.09)	**.57** (.20)	**.61*** **
Lawyerese with Negative Feedback N=20	**9.50** (1.79)	**4.81** (.76)	**.55** (.28)	**.29**

Note: Standard deviations are shown in brackets; ***p<.001

However, as predicted, one-way ANOVAs on the 'difficult' items showed a significant main effect for accuracy, F(2,57) = 4.54, p = .02. The means and standard deviations are shown in Table 8.3.4. Tukey-HSD post-hoc tests (p<.05) found that the mean accuracy score for the Lawyerese with Negative Feedback condition was significantly lower than that shown in the Control condition; none of the other comparisons was significant. A significant main

effect was also found for the within subjects confidence-accuracy correlations, $F(2,57) = 6.86$, $p = .01$. Tukey tests on mean within subjects C-A correlations (see also Table 8.3.4) showed that the C-A relationship in the standard Lawyerese condition was significantly poorer (it was in fact negative) than in both the Control and Lawyerese with Negative Feedback conditions, which did not differ significantly from each other. No significant main effect was found for confidence in replies to the difficult questions. Analyses of between subjects C-A relationships found no significant relationships.

Table 8.3.4 Mean accuracy, confidence, C-A (w-s), and C-A (b-s) correlations for 'Difficult' questions by questioning style

STYLE	Accuracy	Confidence	C-A (w-s)	C-A (b-s)
Control N=20	**5.05** (.89)	**5.01** (1.21)	**.06** (.41)	**.02**
Lawyerese N=20	**4.85** (1.79)	**4.96** (.87)	**-.37** (.32)	**-.26**
Lawyerese with Negative Feedback N=20	**3.65** (1.90)	**4.67** (.94)	**-.08** (.40)	**-.10**

Note: Standard deviations are shown in brackets

In sum, with regard to difficult questions, the two Lawyerese conditions had somewhat different effects. The straight Lawyerese questioning did not reduce the number of correct responses, however, it had a detrimental effect on

the within subjects confidence-accuracy relationships. In contrast the Lawyerese with Negative Feedback questioning had a detrimental effect on accuracy, but not on confidence-accuracy relationships.

In order to further investigate the relationships between confidence and accuracy, mean confidence in 'incorrect' answers was compared to mean confidence in 'correct' answers for each condition (i.e. Control, Lawyerese, and Lawyerese with Negative Feedback) for the 'overall' data (i.e. all items) by means of a 3 x 2 ANOVA with repeated measures on the second factor (Control; Lawyerese; Lawyerese with Negative Feedback X Confidence in Incorrect; Confidence in Correct). The means and standard deviations are shown in Table 8.3.5. No main effects between groups were found (p>.1). However, a significant main effect within subjects for confidence was observed, $F(1,57) = 422.53$, $p = .000$; confidence was higher for 'correct' answers (M = 6.48, S.D. = .78) than 'incorrect' answers (M = 4.28, S.D. = .88). No other significant effects were observed.

Table 8.3.5 Mean confidence in correct and incorrect answers (including overall means) to 'All' questions by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
5.38 (.72)	**6.48** (.78)	**4.28** (.88)
Control N=20 **5.49** (.83)	**6.55** (.88)	**4.43** (.95)
Lawyerese N=20 **5.46** (.72)	**6.50** (.76)	**4.42** (.92)
Lawyerese with Negative Feedback N=20 **5.19** (.60)	**6.38** (.72)	**4.00** (.73)

Note: Standard deviations are shown in brackets

The mean confidence in 'correct' answers was then compared with mean confidence in 'incorrect' answers for each condition (i.e. Control, Lawyerese, and Lawyerese with Negative Feedback) for 'Easy' items again by a 3 x 2 mixed ANOVA with repeated measures on the second factor (Control; Lawyerese; Negative feedback X Confidence in Correct; Confidence in Incorrect). The means and standard deviations are shown in Table 8.3.6.

A significant main effect for groups was observed, $F(2,57) = 5.43$, $p = .01$; confidence was highest in the Control condition and lowest in the

Lawyerese with Negative Feedback condition. Tukey-HSD post-hoc tests ($p<.05$) found that the mean confidence score for the Lawyerese with Negative Feedback was significantly lower (M = 5.22, S.D. = .97) than that shown in the Control condition (M = 6.31, S.D. =1.09); but neither differed from the mean of the Lawyerese condition (M = 5.59, S.D. = .82).

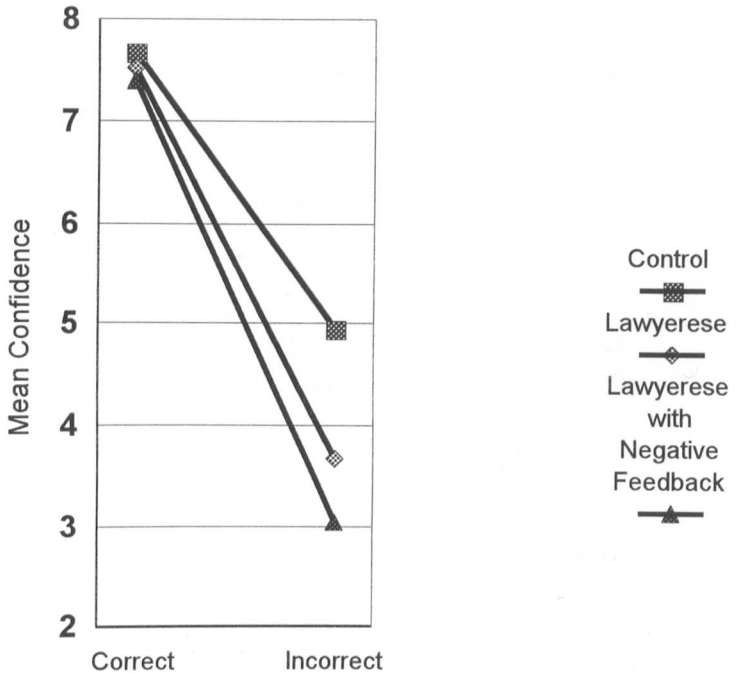
A significant main effect within subjects was also observed, F(1.57) = 203.13, p = .001; confidence was highest for 'correct' answers (M = 7.53, S.D. = .75) and lowest for 'incorrect' answers (M = 3.88, S.D. = 2.04). A significant interaction was also observed, F(2,57) = 3.55, p = .05, see Figure 8.3.6.1. Tukey-HSD post-hoc tests ($p<.05$) found that confidence in 'incorrect' answers was significantly lower in the Lawyerese with Negative Feedback condition (M = 3.04, S.D. = 1.93) than in the Control condition (M = 4.94, S.D. = 1.86); but neither differed from the mean of the Lawyerese condition (M = 3.66, S.D. = 1.94). No significant effects were observed for similar comparisons with regard to confidence in 'correct' answers. Further F tests showed that, within the Control condition, the Lawyerese condition and the Lawyerese with Negative Feedback condition, confidence in 'correct' answers was significantly higher than the confidence shown in 'incorrect' answers ($p<.001$). In other words, the significant positive correlation between subjects confidence-accuracy correlation for 'easy' items with respect to the Lawyerese with Negative Feedback condition (see Table 8.3.2) was due primarily to a decrease in confidence in 'incorrect' answers; a trend also apparent, but not significant in the Lawyerese condition.

Table 8.3.6 Mean confidence in correct and incorrect answers (including overall means) to 'Easy' questions by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
5.71 (1.14)	**7.53** (.75)	**3.88** (2.04)
Control		
N=20 **6.31** (1.09)	**7.68** (.72)	**4.94** (1.86)
Lawyerese		
N=20 **5.59** (.82)	**7.52** (.86)	**3.66** (1.94)
Lawyerese with Negative Feedback N=20 **5.22** (.97)	**7.40** (.68)	**3.04** (1.93)

Note: Standard deviations are shown in brackets

Figure 8.3.6 **Witness confidence ratings in correct and incorrect answers to 'Easy' items for each questioning style**



Mean confidence in 'incorrect' answers was then compared with mean confidence in 'correct' answers for each condition (i.e. Control, Lawyerese, and Negative feedback), this time for 'Moderate' items by a 3 x 2 ANOVA (Control; Lawyerese; Negative feedback X Confidence in Incorrect; Confidence in Correct, with repeated measures on the second factor). No main effects between groups were found (p>.1). However, a significant main effect within subjects for confidence was observed, F(1,57) = 192.15, p = .001; confidence was higher in 'correct' answers (M = 6.15, S.D. = 1.15) than 'incorrect' answers (M = 3.21, S.D. = 1.17). The means and standard deviations are shown in Table 8.3.7. No significant interaction was present.

Table 8.3.7 Mean confidence in correct and incorrect answers (including overall means) to 'Moderate' questions by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
4.68 (.82)	**6.15** (1.15)	**3.21** (1.17)
Control N=20 **4.69** (.87)	**5.97** (1.26)	**3.41** (1.18)
Lawyerese N=20 **4.92** (.87)	**6.54** (1.09)	**3.29** (1.23)
Lawyerese with Negative Feedback N=20 **4.42** (.68)	**5.92** (1.03)	**2.93** (1.15)

Note: Standard deviations are shown in brackets

Mean confidence in 'correct' answers was then compared with mean confidence in 'incorrect' answers for each condition (i.e. Control, Lawyerese, and Lawyerese with Negative Feedback), for 'Difficult' items by means of a 3 x 2 mixed ANOVA (Control; Lawyerese; Negative feedback X Confidence in Correct; Confidence in Incorrect) with repeated measures on the second factor. The means and standard deviations are shown in Table 8.3.8.

No main effects between groups were found. However, a significant main effect within subjects for confidence was observed $F(1,57) = 6.80$, $p = .02$; confidence was higher for 'incorrect' answers ($M = 5.25$, $SD = 1.27$) than

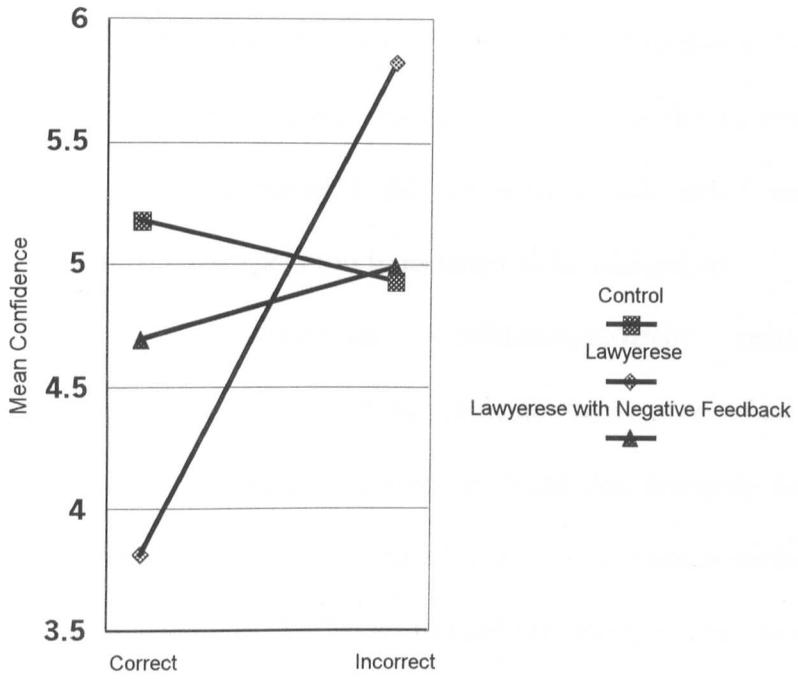
'correct' answers (M = 4.56, SD = 1.80). A significant interaction was also observed, F(2,57) = 6.70, p = .01; see Figure 8.3.8.1. Tukey-HSD post-hoc tests (p<.05) found that confidence in 'correct' answers was significantly lower in the standard Lawyerese condition (M = 3.81, SD = 1.28, p = .05) than in either the Control (M = 5.18, SD = 1.64) or Lawyerese with Negative Feedback (M = 4.69, SD = 2.17) conditions, although the latter two did not differ from each other. No significant effects were observed for similar comparisons with regard to confidence in 'incorrect' responses. Further F tests showed that within the Lawyerese condition confidence in 'correct' answers was significantly lower than confidence in 'incorrect' answers (p<.001). In other words, the negative within subjects confidence-accuracy correlation for the straight Lawyerese condition (see Table 8.3.4) was due primarily to a decrease in confidence in 'correct' answers; a trend also apparent but not significant in the Lawyerese with Negative Feedback condition. None of the other comparisons was significant.

Table 8.3.8 Mean confidence in correct and incorrect answers (including overall means) to 'Difficult' questions by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
4.91 (1.09)	**4.56** (1.80)	**5.25** (1.27)
Control		
N=20 **5.06** (1.22)	**5.18** (1.64)	**4.93** (1.46)
Lawyerese		
N=20 **4.82** (.82)	**3.81** (1.28)	**5.82** (1.23)
Lawyerese with Negative Feedback		
N=20 **4.84** (1.22)	**4.69** (2.17)	**4.99** (.91)

Note: Standard deviations are shown in brackets

Figure 8.3.8 **Witness confidence ratings in correct and incorrect answers to 'Difficult' items for each questioning style**



8.4 DISCUSSION

This study showed that witnesses were significantly less accurate in their responses to the 'difficult' items when subtle feedback was present during the Lawyerese interview, as compared with a simple Control interview. One possible explanation for this is that the feedback has a disconcerting effect upon witnesses leading them to some confusion, thus reducing ability to adequately check their memories through the detection of discrepancies (Schooler & Loftus, 1986). It may be the case, therefore, that, despite a variety of results suggesting that leading questions may impede witness accuracy (Dent, 1978; Dent & Stephenson, 1979; Kebbell & Giles, 2000; Lipton, 1977; Loftus, 1975; Marquis, Marshall & Oskamp, 1972; Turtle &

Wells, 1988), in a setting more analogous to the courtroom, witness' accuracy is only significantly affected when the questions are accompanied by negative feedback. It is important to note here, however, that these results are not necessarily generalisable to more overt forms of feedback. In experiment 1, for example, overt negative feedback (i.e. 'try harder to remember?'; 'think about that answer again?') did not detrimentally affect accuracy, possibly because such feedback is too transparent to be misleading.

However, witnesses' confidence-accuracy relationships were significantly reduced when these 'difficult' items were asked in the standard Lawyerese style alone; moreover, this was due primarily to a reduction in confidence in correct answers for this condition. One possible explanation for this is that leading the witnesses toward the wrong answer could have resulted in them experiencing feelings of uncertainty in 'correct' answers. However, this uncertainty was not sufficiently confusing to affect accuracy.

Surprisingly, however, no comparable reduction in this confidence-accuracy relationship was shown when subtle feedback was presented that questioned accuracy. Indeed, for the combined data, and the data for the easy items, there was actually a significant positive between subjects confidence-accuracy correlation for this condition. The explanation for these effects may be somewhat complex. It will be remembered that the study was designed such that, in answer to difficult questions, the correct answer always required a 'no' response. In the subtle feedback condition the Lawyerese questioning might have first served to reduce confidence in correct answers. However, when a witness said 'no' to any question, subtle feedback was then applied to attempt to shift the response to a 'yes' answer. This might have given the

impression to the witness that whenever he or she gave 'no' as an answer, this would be challenged. As a result, the witness might only have said 'no' when he or she was relatively certain that response was correct (which it was). In other words, the negative effects of the Lawyerese questioning on confidence in correct responses might have been cancelled out by the challenge of subtle feedback, which reduced the number of correct responses but increased confidence in those fewer answers that were correct. As a result, the confidence-accuracy correlations in this condition were not as poor as when Lawyerese questioning alone was used. A general tendency for witnesses to be more strict about applying a high confidence to an answer in the negative feedback condition might also help to explain the significant overall (i.e. all item types combined) and easy items positive between subjects confidence-accuracy correlations for this condition (the former was clearly due to the latter). When the items were easy, participants maintained their confidence in correct answers despite the negative feedback (because the answers were obvious) but decreased their confidence in those answers that were incorrect (less obvious) and therefore may have been more amenable to influence by negative feedback.

CHAPTER 9

CROSS-EXAMINATION QUESTIONING STYLES: Experiment 3

9.1 INTRODUCTION

As mentioned earlier, in order to make experiment 2 more ecologically valid, correct responses to 'difficult' responses required a 'no' response. Nevertheless, it is of interest to know whether significant results occurred only in the 'difficult' item condition solely because a 'no' response was required rather than because they were difficult. Accordingly, a follow up study was conducted using a modified schedule to test the effects of 'nay-saying' per se.

9.2 METHOD

9.2.1 Participants

The participants were 11 males and 19 females drawn from an opportunity sample. All participants were students at the University of Liverpool. The population mean age was 22.37 (SD = 3.74; range = 19-37). Again, participants were randomly assigned to one of the three conditions; 'Control' (N=10), 'Lawyerese' (N=10) or 'Lawyerese with Negative Feedback' (N=10).

9.2.2 Materials and procedure

The materials and procedure were identical to experiment 2 except that in the interview schedule the 10 difficult questions were substituted with 10 additional easy questions that required a 'no' answer to be correct. In other words, there were 38 items, 14 moderate and 24 easy (of which 10 required a

'no' answer; see Appendices 8, 9, and 10). If the effects were solely due to saying 'no', or influenced by this factor, they should occur primarily for the easy items requiring a 'no' answer.

9.3 RESULTS

One-way ANOVAs (Control, Lawyerese, Lawyerese with Negative Feedback) were performed on the 'overall' data for accuracy (mean number of items correct), confidence (mean scores on the 1-9 Likert scale), and within subjects confidence-accuracy correlations (C-A; mean of the confidence accuracy correlations for each participant). Again, all correlations are Pearsons unless otherwise specified. None of the main effects was significant for any of these analyses. The means and standard deviations are shown in Table 9.3.1. However, a trend for accuracy was observed $F(2,27) = 2.63$, $p = .090$; accuracy was highest in the Lawyerese and lowest in the Lawyerese with Negative Feedback condition. A trend was also observed for confidence, $F(2.27) = 2.67$, $p = .088$; confidence was highest in the Lawyerese condition and lowest in the Control condition.

The only significant effects to emerge from the combined data were significant between subjects' confidence-accuracy correlations (i.e. the correlation between the total accuracy score and the total confidence score for each witness) for the Control condition ($r=.82$, n=10, $p< .01$) and the Lawyerese with Negative feedback condition ($r=.72$, n=10, $p< .02$); the same correlation for the Lawyerese condition was positive but not significant (.30); though the correlations were not significantly different from each other (z statistic, $p>.05$).

Table 9.3.1 Mean accuracy, confidence, C-A (w-s), and C-A (b-s) correlations for 'All' questions by style of questioning

STYLE	Accuracy	Confidence	C-A (w-s)	C-A (b-s)
Control N=10	**30.90** (2.47)	**5.95** (1.15)	**.49** (.13)	**.82****
Lawyerese N=10	**32.70** (2.63)	**6.92** (.73)	**.41** (.16)	**.30**
Lawyerese with Negative Feedback N=10	**30.20** (2.44)	**6.49** (.89)	**.47** (.13)	**.72***

Note: Standard deviations are shown in brackets; **p<.01; *p<.02

One-way ANOVAs were then performed on the 10 'easy' target items in each condition (Control, Lawyerese, Lawyerese with Negative Feedback) where a 'no' response was required to be accurate. Separate analyses were conducted for mean accuracy, mean confidence, and within subjects confidence-accuracy correlations. Between subjects C-A relationships were again calculated for each group. No significant effects were found, although a trend was observed for accuracy $F(2,27) = 2.61$, $p=.092$; highest accuracy was shown in the Lawyerese condition. A trend was also observed for the confidence-accuracy correlation in the Lawyerese with Negative feedback condition ($r=.59$, n=10, $p = .070$). It can also be noted that there was a quite pronounced non-significant negative between subjects confidence-accuracy correlation for the Lawyerese group (-.48). The means and standard deviations are shown in Table 9.3.2.

Table 9.3.2 Mean accuracy, confidence, C-A (w-s), and C-A (b-s) correlations for 'Easy' questions by style of questioning

STYLE	Accuracy	Confidence	C-A (w-s)	C-A (b-s)
Control N=10	**8.50** (.97)	**5.78** (1.13)	**.30** (.32)	**.29**
Lawyerese N=10	**9.10** (.88)	**6.38** (.95)	**.21** (.32)	**-.48**
Lawyerese with Negative Feedback N=10	**8.00** (1.33)	**6.00** (1.07)	**.37** (.28)	**.59**

Note: Standard deviations are shown in brackets

In order to investigate further the relationships between confidence and accuracy, mean confidence in 'incorrect' answers was compared to mean confidence in 'correct' answers for each condition (i.e. Control, Lawyerese, and Lawyerese with Negative Feedback) for the 'overall' data (i.e. all items) by means of a 3 x 2 ANOVA (Control; Lawyerese; Lawyerese with Negative Feedback X Confidence in Incorrect; Confidence in Correct) with repeated measures on the second factor. The means and standard deviations are shown in Table 9.3.3. No main effects between groups were found (p>.1). However, a significant main effect within subjects for confidence was observed, F(1,27) = 207.88, p = .001; confidence was higher in 'correct' answers (M = 7.01, S.D. = .89) than 'incorrect' answers (M = 3.96, S.D. = 1.40). No significant interaction was present.

Table 9.3.3 Mean confidence in correct and incorrect answers (including overall means) to 'All' questions by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
5.49 (1.03)	**7.01** (.89)	**3.96** (1.40)
Control N=20 **4.99** (1.01)	**6.60** (1.16)	**3.39** (1.07)
Lawyerese N=20 **5.86** (1.07)	**7.32** (.66)	**4.40** (1.73)
Lawyerese with Negative Feedback N=20 **5.61** (.91)	**7.12** (.69)	**4.09** (1.26)

Note: Standard deviations are shown in brackets

Mean confidence in 'incorrect' answers was again compared to mean confidence in 'correct' answers for each condition (i.e. Control, Lawyerese, and Lawyerese with Negative Feedback) but this time for the 10 substituted items for which a 'no response' was required using a 3 x 2 ANOVA with repeated measures on the second factor (Control; Lawyerese; Lawyerese with Negative Feedback X Confidence in Incorrect; Confidence in Correct). The means and standard deviations are shown in Table 9.3.4. No main effects between groups were found (p>.1). However, again a significant main effect within subjects for confidence was observed, $F(1,27) = 62.50$, $p = .001$;

confidence was higher in 'correct' answers (M = 6.43, S.D. = 1.09) than 'incorrect' answers (M = 3.30, S.D. = 2.36). No significant interaction was present.

Table 9.3.4 Mean confidence in correct and incorrect answers (including overall means) to 'Easy' questions by questioning style

STYLE		
Overall	**Confidence/Correct**	**Confidence/Incorrect**
4.86 (1.47)	**6.43** (1.09)	**3.30** (2.36)
Control N=20 **4.50** (1.29)	**6.10** (1.20)	**2.90** (2.13)
Lawyerese N=20 **4.68** (1.77)	**6.66** (1.21)	**2.70** (2.83)
Lawyerese with Negative Feedback N=20 **5.41** (1.26)	**6.53** (.84)	**4.29** (1.92)

Note: Standard deviations are shown in brackets

9.4 DISCUSSION

The negative between subjects confidence-accuracy effect for the lawyerese condition (see Table 9.3.2) may be worthy of further investigation on a larger sample, but on the whole, these data suggest the significant effects found for

the 'difficult' questions occurred primarily because of the nature of the material in the question rather than because a 'no' response was necessary for a correct answer.

Interesting additional findings in this study were the impressive between subjects correlations for the control and lawyerese with negative feedback conditions shown also in the analyses of confidence in incorrect and incorrect items. This fits with other recent findings suggesting that confidence-accuracy relationships may not necessarily be as poor as early researchers suggested (Kebbell, Wagstaff & Covey, 1996). The fact that correlations of similar magnitude were not found in the previous study would seem to have occurred because of the deleterious effects of difficult items.

CHAPTER 10

DELAY AND THE STABILITY OF CONFIDENCE AND ACCURACY: Experiment 4

10.1 INTRODUCTION

Recently the Home Office has produced a protocol in an attempt to reduce delays in the Youth Justice System (see Home Office, 1999; Glidewell Recommendations 21 and 22 in the Magistrates' Courts). This reflects a general desire to reduce delays in the Criminal Justice system, including the time between initial police questioning and subsequent appearances in court, and also the time between initial appearances in court and appeals. Given this emphasis, research on the effects of delay on eyewitnesses' accuracy, confidence, and the relationship between confidence and accuracy, would seem to be essential.

Few studies have examined the effects of delay. Two studies have shown that repeat questioning of the same material can increase confidence after 2 days (Hastie, Landsman & Loftus, 1978), and after 3 weeks (Turtle & Yuille, 1994). In contrast, however, Ryan and Geiselman, (1991) reported that confidence decreased, following repeat questioning after 1 week, between the test and retest situation, and more so for incorrect than correct answers; the result was an increase in confidence-accuracy judgements over time. However, a study by Granhag (1997) suggests that the positive effect of delay (1 week) on confidence-accuracy may be more a function of making repeated confidence ratings *per se*, than the delay involved. Thus far, therefore, the results on repeated recall on confidence are contradictory, and no work has

examined these effects over a time period more akin to that experienced by witnesses in the Criminal Justice System, or examined the effects of item difficulty.

While it might have been beneficial to study longitudinally all three conditions used in experiment 2, because of time pressures and participant availability it was decided to concentrate on the standard lawyerese condition only. The standard lawyerese condition was selected for further study, mainly because it mirrors more closely cross-examination as might be performed in the courtroom. Also, of the three examination styles, the lawyerese interview had the most detrimental effect on confidence-accuracy relationships. Consequently, in the following experiment, a new sample of participants was presented with the same materials as those in experiment 2 and tested after five minutes and six months. Additionally, because a different sample of participants was used in the following study it was possible to compare the results with those of experiment 2.

10.1.1 Hypotheses

Intuitively, given standard theories of memory decay, one might expect that both confidence and accuracy would reduce over a long delay. However, an alternative hypothesis is that memory might decay selectively; for example, data from this thesis and previous studies (Kebbell, Wagstaff, & Covey, 1996) indicate that when participants are absolutely sure of a response they tend to be accurate. Arguably, such responses are the least likely to decay over time, not only because of the strength of the original association, but also because they might be the most obvious candidates for rehearsal. In contrast, other

responses that have weaker associations might be more likely to decay over time, with a corresponding decline in confidence. One result of this might be a decline in the overall number of correct responses over time, but an increase in confidence-accuracy relationships as the distinction between the two types of response (accurate sure, inaccurate unsure) becomes more exaggerated. This would account for the finding by Ryan and Geiselman (1991) that confidence-accuracy increased after a delay largely due to the lowered confidence in incorrect answers. It should be noted that Granhag (1997) found that this effect was only evident (after one week) for those who had previously rated their confidence in material and had an opportunity to review their previous answers (repeated rating). Given that, to accord with ecological validity, participants were not formally requested to review their previous confidence ratings, the hypotheses for the following experiment were not directional at this stage.

10.2 METHOD

10.2.1 Participants

The participants were 4 males and 17 females drawn from an opportunity sample. The mean age of the sample was 20.95 (SD = 5.45; range = 18-44). All participants were either students at the University of Liverpool or members of a research panel in the Department of Psychology at the University.

10.2.2 Materials and procedure

The participants for longitudinal study were first required to undergo a neurological test of verbal fluency (i.e. phonemic). The purpose and results of this test are reported in Chapter 12. The materials and procedure used following this test was identical to those previously described for the standard Lawyerese condition (see Chapter 8).

As before, all participants were told that they were to observe a videotape of an event sequence for a period of around 4-5 minutes, and afterwards they would be required to answer some questions. All participants were then shown the five minutes colour video depicting a criminal offence.

After this, every participant was required to complete a filler task for a period of five minutes, which involved reading unrelated material. Participants were advised that they would now be asked a series of questions regarding the videotape they had observed. All questions required 'yes' or 'no' answers, and following each question participants were asked to rate their confidence in the response they had given on a Likert scale ranging from 1 to 9, where (1) represented 'pure guess' and (9) represented 'absolutely certain'.

Each participant was then individually interviewed under the standard Lawyerese condition. As noted previously, in the Lawyerese condition participants were required to give answers to the three sets of target items (i.e. 14 'easy', 14 'moderate' and 10 'difficult' items), but the questions were so phrased as to replicate the manner in which lawyers conduct cross-examinations in court.

At the end of the interview schedule, all participants completed a series of questionnaires (also described and reported in Chapter 12), were thanked

for their participation and debriefed. Again, the rationale for the experiment was explained to each participant that created opportunities for participants to ask questions and raise issues for discussion. These are highlighted on p. 189 of the discussion.

The participants were then required to return for re-interview six months later.

10.3 RESULTS

10.3.1 Within-subjects analysis (5 minute and 6 month delay)

Accuracy (mean number of items correct); confidence (mean scores on the 1-9 Likert scale; where 1 represented 'pure guess' and 9 represented 'absolutely certain'); and within subjects confidence-accuracy correlations (mean of the confidence-accuracy correlations for each participant) were analysed for the overall data (easy, moderate and difficult items combined) by related t-tests (after 5 minutes and 6 months). Once more, all correlations are Pearsons unless otherwise specified. Both accuracy and confidence were found to be significantly lower at 6 months ($t = 6.67$, $df = 20$, $p = .001$ and $t = 10.15$, $df = 20$, $p = .001$, respectively), see Table 10.3.1. No significant effects were found for within subjects C-A relationships.

The data for the same dependent variables were then analysed for the 'easy' and 'moderate' items by means of three 2 x 2 repeated measures ANOVAs (after 5 minutes/after 6 months X 'easy'/'moderate' items). The means and standard deviations are also shown in Table 10.3.1. Difficult items

were not included in this analysis for reasons stated earlier (i.e. there was a different number (10) of them, and they all required a 'no' response). A significant main effect for accuracy was observed for item difficulty, $F(1,20) = 31.25$, $p<.001$; accuracy was higher for easy items. A significant main effect was also found for retention interval, $F(1,20) = 16.28$, $p<.001$; accuracy was lower at 6 months.

A significant main effect for confidence was also observed for item difficulty, $F(1,20) = 126.29$, $p<.001$; confidence was also highest for easy items. A significant main effect for confidence was also observed for retention interval, $F(1,20) = 131.48$, $p<.001$; lowest confidence was also expressed at 6 months. No significant effects were found for within subjects C-A correlations and no significant interaction was present.

Accuracy, confidence, and within subjects confidence-accuracy relationships were then analysed for the 10 'difficult' items by related t-tests (comparing results after 5 minutes and 6 months; see again Table 10.3.1). In keeping with the data for the easy and moderate items, both accuracy and confidence for 'difficult' items were found to be significantly lower 6 months later ($t = 3.48$, $df = 20$, $p<.002$ and $t = 5.77$, $df = 20$, $p< .001$, respectively). No effects for within subjects C-A were found.

Between subjects C-A relationships were also calculated 'overall' for each retention interval (5 minutes; 6 months) and for item difficulty (easy; moderate; difficult; see also Table 10.3.1). Between subjects correlations for the 'overall' data; 'easy', and 'moderate' items were all found to be significant at 6 months ($r=.52$, $p<.02$; $r=.50$, $p<.05$; $r=.52$, $p<.02$, respectively) whereas the between subjects correlation for the 'difficult' items was not ($r=-.38$, n.s.).

While the relationships for the easy and moderate items did not differ significantly from each other (z statistic, $p>.05$) both were found to differ significantly from the correlation calculated for the 'difficult' items (z statistic, $p<.05$). No significant correlations were found for the 5 minutes delay.

Table 10.3.1 Within-subjects mean Accuracy, Confidence, C-A (w-s), and C-A (b-s) for Overall, Easy, Moderate, and Difficult items at 5 minutes and 6 months

	5 Minute Delay (N=21)				**6 Month Delay** (N=21)			
	All	**Easy**	**Mod**	**Diff**	**All**	**Easy**	**Mod**	**Diff**
Accuracy	**27.95** (2.16)	**12.95** (.92)	**10.38** (1.47)	**4.62** (1.94)	**23.48** (2.79)	**11.29** (1.87)	**9.19** (3.03)	**3.00** (2.24)
Conf	**6.29** (1.00)	**7.49** (.98)	**5.67** (1.26)	**5.49** (1.10)	**4.55** (1.25)	**5.50** (1.32)	**3.92** (1.29)	**4.09** (1.49)
C-A (w-s)	**.35** (.14)	**.40** (.33)	**.47** (.22)	**-.11** (.33)	**.31** (.15)	**.43** (.25)	**.40** (.32)	**-.19** (.37)
C-A (b-s)	**.12**	**.28**	**.21**	**-.12**	**.52***	**.50***	**.52***	**-.38**

Note: Standard deviations are shown in brackets; *$p<.05$

In sum, as expected, the delay from giving verbal report at 5 minutes to 6 months was found to significantly reduce both accuracy and confidence. However, while within subjects C-A correlations were unaffected by this delay, between subjects C-A relationships became significant in a positive direction at six months for the interview overall, for easy items, and for moderate items. In contrast, both within and between subjects C-A

relationships for the difficult items remained negative and statistically insignificant over the 6 months interval.

10.3.2 Between-subjects analysis (5 minute and 6 month delay)

For reliability purposes, it was considered useful to compare the results of this study with those of the lawyerese condition from experiment 2. In experiment 2 participants were given exactly the same materials as here, but tested after 5 minutes only. As a first analysis, therefore, accuracy (mean number of items correct); confidence (mean scores on the 1-9 Likert scale); and within subjects confidence-accuracy correlations (C-A; mean of the confidence-accuracy correlations for each participant) were analysed for the 'overall' data (easy, moderate and difficult items combined) by unrelated t-tests comparing the two samples, experiment 2 and here, after a 5 minutes delay and 6 months delay respectively.

As before, both accuracy and confidence were found to be significantly lower at 6 months (t = 2.76, df = 38, p<.01 and t = 3.83, df = 38, p<.001, respectively), see Table 10.3.2. No significant effects were found for within subjects C-A relationships.

Again the data for the same dependent variables were then analysed for the 'easy' and 'moderate' items by means of three 2 x 2 mixed ANOVAs (after 5 minutes/after 6 months X 'easy'/'moderate' items). The means and standard deviations are also shown in Table 10.3.2.

Again a significant main effect for accuracy was observed for item difficulty, F(1,38) = 32.68, p<.001; accuracy was higher for easy items.

A significant main effect for confidence was also found for item difficulty, $F(1,38) = 117.17$, $p<.001$; confidence was higher for easy items. A significant main effect for confidence was also observed for retention interval, $F(1,38) = 18.21$, $p<.001$; confidence was lower at 6 months. Unlike previously, a significant main effect for retention interval was also found for within subjects C-A correlations, $F(1,38) = 6.07$, $p<.02$; the largest positive relationship was observed at 5 minutes. No other significant effects were observed and no significant interaction was present.

Again, accuracy, confidence, and within subjects confidence-accuracy relationships were then analysed for the 10 'difficult' items by unrelated t-tests (comparing results again after 5 minutes and 6 months; see again Table 10.3.2). Again, both accuracy and confidence were found to be significantly lower at 6 months ($t = 2.73$, $df = 38$, $p<.01$ and $t = 2.13$, $df = 38$, $p<.05$, respectively). No effects for within subjects C-A were found.

In sum, with the exception of one result (between subjects C-A for moderate items) the results observed here tend to mirror those found in Table 10.3.1; lending some support for the reliability of the interview procedure.

Table 10.3.2 Between-subjects mean Accuracy, Confidence, C-A (w-s), and C-A (b-s) for Overall, Easy, Moderate, and Difficult items at 5 minutes and 6 months

		Accuracy	Confidence	C-A (w-s)	C-A (b-s)
	Overall	**26.05** (2.76)	**5.86** (.75)	**.35** (.12)	**.38**
5 Minute Delay	**Easy**	**12.10** (1.02)	**6.97** (.81)	**.54** (.26)	**.33**
(N=20)	**Moderate**	**9.10** (2.25)	**5.39** (1.09)	**.57** (.20)	**.61***
	Difficult	**4.85** (1.79)	**4.96** (.87)	**-.37** (.32)	**-.26**
	Overall	**23.65** (2.74)	**4.59** (1.27)	**.31** (.15)	**.51***
6 Month Delay	**Easy**	**11.25** (1.92)	**5.52** (1.35)	**.43** (.26)	**.51***
(N=20)	**Moderate**	**9.30** (3.06)	**3.97** (1.30)	**.38** (.32)	**.51***
	Difficult	**3.10** (2.25)	**4.13** (1.51)	**-.20** (.37)	**-.42**

Note: Standard deviations are shown in brackets; *$p<.05$

10.3.3 Confidence in Correct and Incorrect Answers

Returning to the data produced solely by participants in the present experiment, mean confidence in 'incorrect' answers was compared to mean confidence in 'correct' answers for each retention interval (5 Minutes and 6 Months) for the 'overall' data (i.e. easy, moderate, and difficult items combined) by means of a 2 x 2 repeated measures ANOVA (5 Minutes/6

Months X Confidence in Incorrect/Confidence in Correct). The means and standard deviations are shown in Table 10.3.3. A significant main effect was observed for retention interval, $F(1,20) = 70.80$, $p<.001$; confidence was higher at 5 minutes (M = 5.84, S.D. = 1.47) than at 6 months (M = 4.38, S.D. = 1.4.3). A significant main effect for Incorrect/Correct was also found, $F(1,20) = 139.00$, $p<.001$; confidence was higher for correct answers (M = 5.97, S.D. = 1.42) than incorrect answers (M = 4.25, S.D. = 1.34). No significant interaction was present, thus confidence in correct and incorrect answers reduced similarly over time.

Table 10.3.3 Mean confidence in correct and incorrect answers (including overall means) to 'All' questions by retention interval

INTERVAL	Confidence/Correct	Confidence/Incorrect
	5.97 (1.42)	**4.25** (1.34)
5 Minutes **5.84** (1.47) N=20	**6.80** (1.00)	**4.88** (1.22)
6 Months **4.38** (1.43) N=20	**5.14** (1.29)	**3.63** (1.17)

Note: Standard deviations are shown in brackets

The mean confidence in 'correct' answers was then compared with mean confidence in 'incorrect' answers for each retention interval (5 Minutes and 6

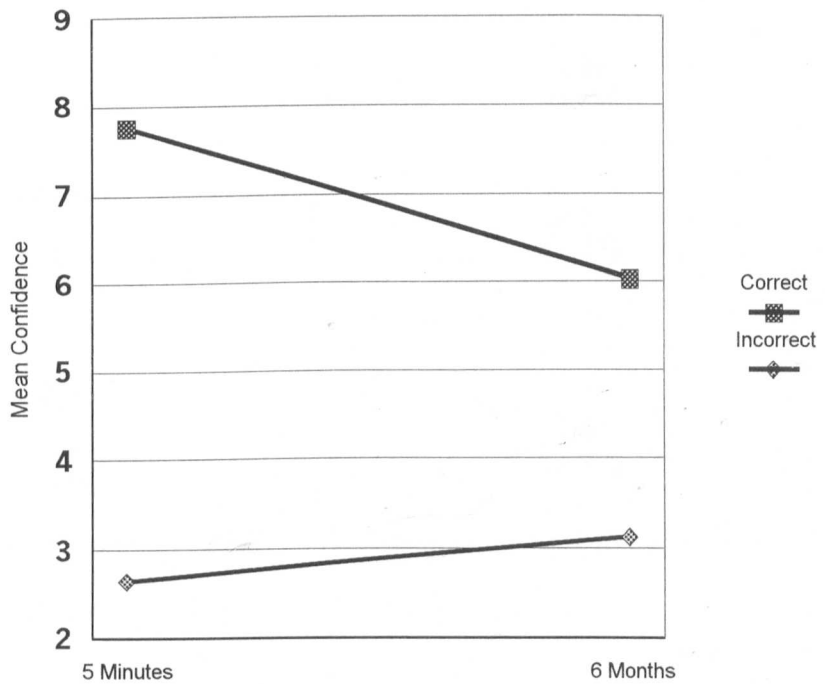
Months) for 'Easy' items again by a 2 x 2 repeated measures ANOVA (5 Minutes/6 Months X Confidence in Correct/Confidence in Incorrect). The means and standard deviations are shown in Table 10.3.4. No significant effects were found between the two intervals. However, a significant main effect for Incorrect/Correct was observed, $F(1,20) = 128.55$, $p<.001$; confidence was again higher for correct answers ($M = 6.90$, S.D. = 1.42) than incorrect answers ($M = 2.89$, S.D. = 2.14). A significant interaction was also present, $F(1,20) = 14.64$, $p<.01$, see Figure 10.3.4. Further F tests showed that confidence in correct answers at 5 minutes was significantly higher than confidence in correct answers at 6 months ($p<.001$), while no difference was found between confidence in incorrect answers at 5 minutes and 6 months ($p>.05$). Also confidence in correct answers differed significantly from confidence in incorrect answers, both at 5 minutes ($p<.001$) and 6 months ($p<.001$).

Table 10.3.4 Mean confidence in correct and incorrect answers (including overall means) to 'Easy' questions by retention interval

INTERVAL	Confidence/Correct	Confidence/Incorrect
	6.90 (1.42)	**2.89** (2.14)
5 Minutes **5.20** (3.23) N=20	**7.76** (.93)	**2.63** (2.59)
6 Months **4.59** (2.05) N=20	**6.04** (1.31)	**3.15** (1.59)

Note: Standard deviations are shown in brackets

Figure 10.3.4 Witness confidence ratings in correct and incorrect answers to 'Easy' items for each retention interval



Mean confidence in 'incorrect' answers was then compared with mean confidence in 'correct' answers for each retention interval (5 Minutes and 6 Months), this time for 'Moderate' items by a 2 x 2 repeated measures ANOVA (5 Minutes/6 Months X Confidence in Incorrect/Confidence in Correct). The means and standard deviations are shown in Table 10.3.5. A significant main effect was observed for retention interval, $F(1,20) = 25.34$, $p<.001$; confidence was higher at 5 minutes (M = 4.98, S.D. = 2.07) than at 6 months (M = 3.52, S.D. = 1.76). A significant main effect for Incorrect/Correct was also found, $F(1,20) = 91.95$, $p<.001$; confidence was higher for correct answers (M = 5 .51, S.D. = 1.55) than for incorrect answers (M = 2.99, S.D. = 1.68). No significant interaction was observed.

Table 10.3.5 Mean confidence in correct and incorrect answers (including overall means) to 'Moderate' questions by retention interval

INTERVAL	Confidence/Correct	Confidence/Incorrect
	5.51 (1.55)	**2.99** (1.68)
5 Minutes **4.98** (2.07) N=20	**6.36** (1.29)	**3.60** (1.77)
6 Months **3.52** (1.76) N=20	**4.66** (1.33)	**2.38** (1.37)

Note: Standard deviations are shown in brackets

Mean confidence in 'correct' answers was then compared with mean confidence in 'incorrect' answers for each retention interval (5 Minutes and 6 Months), but for 'Difficult' items by means of a 2 x 2 repeated measures ANOVA (5 Minutes/6 Months X Confidence in Correct/Confidence in Incorrect). The means and standard deviations are shown in Table 10.3.6. A significant main effect was observed for retention interval, $F(1,20) = 21.45$, $p<.001$; confidence was higher at 5 minutes (M = 5.23, S.D. = 1.65) than at 6 months (M =3.78, S.D. = 2.00). A significant main effect for Incorrect/Correct was also found, $F(1,20) = 8.50$, $p<.01$; confidence was higher for incorrect answers (M = 5.09, S.D. 1.55) than correct ones (M =3.92, S.D. = 2.17). No significant interaction was present.

Table 10.3.6 Mean confidence in correct and incorrect answers (including overall means) to 'Difficult' questions by retention interval

INTERVAL	Confidence/Correct	Confidence/Incorrect
	3.92 (2.17)	**5.09** (1.55)
5 Minutes **5.23** (1.65) N=20	**4.71** (1.83)	**5.76** (1.29)
6 Months **3.78** (2.00) N=20	**3.13** (2.24)	**4.43** (1.52)

Note: Standard deviations are shown in brackets

Because the C-A data for the 5 minutes condition in this experiment and experiment 2 were so similar, between subjects analyses were not warranted in the confidence in correct and confidence in incorrect data.

Finally, out of 798 possible responses given by the 5 minutes interval group, 285 were rated in terms of confidence as "absolutely certain"; of these, 89% were correct. For the 6 months interval group, 103 were rated as "absolutely certain" and, of these, 86% were correct. Overall, 87.5% of "absolutely certain" answers were correct.

10.4 DISCUSSION

To summarize, perhaps not surprisingly, the results showed that a 6 months delay reduced accuracy for all levels of question difficulty. Also, the 6 months delay reduced confidence for all levels of question difficulty except for the easy items. Confidence and accuracy were also highest for easy compared to moderate items; however, these trends were not differentially affected by the delay.

However, whilst within subjects C-A correlations were unaffected by the delay, remaining non-significant, between subjects correlations became significant and positive for the overall data (question difficulty levels combined) and for the easy and moderate questions separately. The between subjects C-A correlation for difficult items after the delay was slightly more negative, though not significantly so.

Although the within subjects C-A correlations were not significant, analyses of confidence in correct and incorrect answers showed, nevertheless, overall, a significant tendency for greater confidence in correct answers for the

easy and moderate items, and in incorrect answers for the difficult items. Overall, however, no differential effects of delay were found, i.e. confidence in correct and incorrect answers declined similarly over time. Thus there was no support for Ryan and Geiselman's (1991) finding that a one week delay led to increased within subjects C-A relationships due to lowered confidence in incorrect answers only, or Granhag's (1997) finding that within subjects C-A correlations increased after a one week interval for participants given an opportunity to review their confidence ratings. However, the most obvious difference between the present study and these others is that a longer time delay was used and, in the absence of a formal review, participants would have been less likely to have remembered their previous responses.

Taken together the results indicate some clear trends. In general, although confidence and accuracy tend to decline over time, C-A relationships tend to remain positive, though not always significant, for easy and moderate items. For difficult items, however, the trend is for non-significant negative C-A relationships, to persist over time. These trends, however, seem to exaggerate over time for between subjects C-A correlations; i.e. the positive relationships improve, whilst the negative relationship becomes worse.

As the trend for the between-subjects correlations to be exaggerated over time was not a consequence of lowered confidence in incorrect answers relative to correct answers, another explanation is called for. Further examination of the data showed more consistent evidence of greater sample variability in scores after the delay. For example, Table 10.3.1 shows larger accuracy and confidence score SDs for all levels of item difficulty after the delay. The trend for greater variability is also shown in the range scores; for

example, for the overall data the range on accuracy after 5 minutes is 8, after 6 months it is 12. The increase in heterogeneity would then allow a higher correlation (Kebbell *et al.*, 1996). Further examination of the raw data showed that this effect was in the main due to participants who were accurate but not confident after 5 minutes becoming inaccurate and not confident after the delay. Why this should happen is unclear; but one possibility is that people who 'guess' are more likely to be correct in their guesses after 5 minutes (because of the availability of partial memory traces). However, their guesses are correspondingly inaccurate after 6 months as the weak traces decay (Weingartner & Parker, 1984). In other words, the initially weak correlations for the easy and moderate items were so much due to participants who were confident in incorrect answers, but because of those who were *not* confident in their correct answers (because they were guessing correctly).

However, the converse would apply if the C-A correlation was initially negative. The effect would be worsened if those who were confident in wrong answers tended to maintain their confidence in these incorrect answers whilst those who were less confident in correct answers tended to forget their correct answers over time.

CHAPTER 11

JURORS' PERCEPTIONS OF WITNESSES: Experiment 5

11.1 INTRODUCTION

The aim of the previous experiments was to assess objectively the effects of questioning styles and techniques used in the courtroom on witness accuracy and confidence; however, given that cross-examination is considered to be central to the English Legal System, perhaps more significant in the context of court decision making processes are the effects of questioning style on the ways in which jurors discriminate between those witnesses who are accurate than those who are not. The aim of the fifth experiment, therefore, was to investigate the effects of questioning style on the inferences made by those hearing the questioning with regard to the accuracy and confidence of the witness being questioned.

One possibility, for instance, is that by suggesting the witness might be wrong, the Lawyerese with Negative Feedback condition might serve to make observers doubt the accuracy of the witness' testimony; for instance, some research has shown that when mock-jurors hear inconsistent recall testimony they perceive the eyewitness to be less accurate and credible (Berman, Narby & Cutler, 1995; Brewer, Potter, Fisher, Bond & Luszcz, 1999). One way inconsistency could operate to lower perceived credibility might be to lower the perceived confidence in the witness. The latter may be particularly important given that, as noted earlier, there is a variety of evidence to indicate that jurors intuitively believe that confidence is a valid predictor of accuracy;

hence many studies have demonstrated that jurors and jurists rely heavily upon the demeanour of the witness; if the witness appears to be confident, he or she will be considered more accurate (Brigham & Wolfskeil, 1983; Cutler & Penrod, 1988; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Stuve, 1988; Deffenbacher & Loftus, 1982; Fox & Walters, 1986; Kassin, Rigby and Castillo, 1991; Leippe, Manion, & Romanczyk, 1992; Lindsay, 1994; Lindsay, Wells & O'Connor, 1989; Lindsay, Wells & Rumpel, 1981; Sporer, 1993; Wells, Ferguson & Lindsay, 1981; Wells, Lindsay & Ferguson, 1979; Whitely & Greenberg, 1986).

Further, in this next experiment, a measure was taken of observers' perceptions of how fair the questioning was considered to be. Although a number of studies have looked at various aspects of procedural fairness in the courtroom, such as adversarial versus non-adversarial procedures (Leung & Lind, 1986; Lind & Tyler, 1988; Wagstaff & Kelhar, 1993), as yet no research has looked at the perceived fairness of different questioning styles within the adversarial system. For instance, jurors might consider that the Lawyerese styles are intimidatory and therefore unfair.

As a matter of interest, additional measures were also taken of observers' perceptions of how stressed the witness seemed to be and how good they thought the witness was.

11.2 METHOD

11.2.1 Participants

The participant observer/jurors were 19 males and 41 females drawn from an

opportunity sample. All participants were students at the University of Liverpool. The mean age of the sample was 22.75 (SD = 7.46; range = 18-58). None had participated in the second experiment.

11.2.2 Materials and procedure

Participants were randomly assigned to one of three conditions; 'Control' (N=20), 'Lawyerese' (N=20) or 'Lawyerese with Negative Feedback' (N=20). Participants in each of the three conditions received audiotapes of voices reading from the transcripts of two witness interviews; one involving a 'good' witness, the other a 'poor' witness. Thus, each participant in the Control, Lawyerese, and Lawyerese with Negative Feedback conditions heard both a 'good' witness interview and one given by a 'poor' witness. The order was counterbalanced such that for half of the participants in each of the three conditions the good witness was first, and for the rest, the poor witness was first.

The witness transcripts were selected on the basis of witnesses' actual accuracy and confidence-accuracy correlations to the interview schedules described in experiment 2. Responses to the whole interview were considered. The criteria for selection were that the witnesses would represent the best and the poorest witness within a condition, compatible with reasonable equivalence over conditions. For the three conditions the accuracy rates and confidence-accuracy relationships of the witnesses selected in this way were as follows: *Control*: 'good' witness accuracy = 30/38; confidence-accuracy relationship for overall interview = .52; and actual mean confidence = 7.32 (SD = 2.05); 'poor' witness accuracy = 25/38; confidence-accuracy

relationship for overall interview = .25; and actual mean confidence = 5.84 (SD = 2.89). *Lawyerese*: 'good' witness accuracy = 31/38; confidence-accuracy relationship for overall interview = .59; and actual mean confidence = 6.68 (SD = 2.77); 'poor' witness accuracy = 24/38; confidence-accuracy relationship for overall interview = .20; and actual mean confidence = 4.95 (SD = 2.69). *Lawyerese with Negative Feedback*: 'good' witness accuracy = 30/38; confidence-accuracy relationship for overall interview = .51; and actual mean confidence = 7.05 (SD = 2.35); 'poor' witness accuracy = 26/38; confidence-accuracy relationship for overall interview = .28; and actual mean confidence = 5.47 (SD = 2.73).

All participants were told that they would be listening to two audio recordings of two separate witnesses being interviewed, one after the other, and that each interview would last for around 3-4 minutes. Participants were also informed that they would be required to complete a questionnaire following each interview. The questionnaire comprised of five questions: a) How accurate do you think the witness was? b) How confident do you think the witness was? c) How stressed do you think the witness was? d) How good a witness do you think the witness actually was? And e) How fair do you think the questioning was? Following each question observers were asked to rate the witness on a Likert scale ranging from 1 to 9, for each. The rating categories were (1) 'extremely inaccurate' to (9) 'extremely accurate'; (1) 'not at all confident' to (9) 'extremely confident'; (1) 'not at all stressed' to (9) 'extremely stressed'; (1) 'extremely poor' to (9) 'extremely good'; and (1) 'not at all fair' to (9) 'extremely fair', for questions a, b, c, d and e, respectively (see Appendix 11).

The audiotapes were standardised across the conditions; two different voices were heard on the tape, one representing the interviewer, the other, the interviewee (eyewitness); the same interviewer asked the questions and the same interviewee gave the responses. The interviewee responded with little intonation. The scripts were based on the exact responses made by two separate witnesses who had observed the same material prior to being interviewed. All participants received the interviews and rated the interviewees either on an individual basis or in groups. For those who participated in pairs or groups no discussion was allowed until the final questionnaire had been completed and collected by the experimenter.

At the end of each session, all participants were thanked for their participation and debriefed fully. At this time, participants received information concerning the nature and rationale for the study and had the opportunity to ask questions and share experiences.

11.3 RESULTS

11.3.1 Accuracy

The observers' accuracy ratings for the two witnesses were analysed by means of a 3 x 2 x 2 mixed ANOVA (Control; Lawyerese; Lawyerese with Negative Feedback X Order; 'good' witness first; 'poor' witness first X good witness; poor witness) with repeated measures on the third factor. The means and standard deviations are shown in Tables 11.3.1.1 and 11.3.1.2. A significant main effect for accuracy was observed between the three interview conditions,

$F(2,54) = 3.51$, $p = .05$. Tukey post-hoc tests ($p<.05$) showed that mean accuracy ratings were significantly lower in the Lawyerese with Negative Feedback condition than in the standard Lawyerese condition (which had the highest mean of all three conditions), but neither differed from the mean of the Control condition. In other words, observers rated the witnesses in the Lawyerese with Negative Feedback condition to be the least accurate.

A significant main effect was also observed for good/poor witness $F(1,54) = 33.39$, $p = .0001$; accuracy ratings were higher for the good witness. A significant interaction was also found (Questioning style condition x good/poor witness), $F(2,54) = 5.44$, $p = .01$, see Figure 11.3.1.1. F tests showed that within both the Control ($p = .002$) and Lawyerese ($p = .001$) conditions, mean accuracy ratings were significantly higher for the good witness than the poor witness, no similar difference was found within the Lawyerese with Negative Feedback condition. Between subjects Tukey-tests ($p<.05$) showed that the good witness was judged to be significantly less accurate in the Lawyerese with Negative Feedback condition. In other words, observers were able to distinguish significantly between the accurate and inaccurate witness in the Control and Lawyerese condition, but not in the Lawyerese with Negative Feedback condition.

The interaction (Order x good/poor witness) was also significant, $F(1,54) = 12.90$, $p = .001$, see Figure 11.3.1.2. F-tests showed that, within subjects, the good witness was rated significantly higher in accuracy than the poor witness, when the good witness was presented first ($p<.01$). However, no difference was observed between the good and poor witnesses' accuracy ratings when the poor witness was presented first ($p>.10$). However, further

F-tests showed that, between subjects, the poor witness was rated as significantly higher in accuracy when presented first rather than second ($p<.02$) while no similar difference existed between the good witness' accuracy ratings. No other significant effects or interactions were observed. In other words, if the good witness is presented first, the good witness is rated to be more accurate than the poor witness; if however, the poor witness is presented first, the perceived accuracy of the poor witness increases such that the difference disappears.

Table 11.3.1.1 Mean Accuracy ratings (including overall means) for taped witness interviews (Good/Poor witness) by questioning style

STYLE	WITNESS	
Overall 5.38 (1.19)	**Good** **6.00** (1.71)	**Poor** **4.77** (1.33)
Control N=20 **5.50** (.89)	**6.20** (1.36)	**4.80** (1.10)
Lawyerese N=20 **5.80** (1.24)	**6.80** (1.44)	**4.80** (1.44)
Lawyerese with Negative Feedback N=20 **4.85** (1.27)	**5.00** (1.84)	**4.70** (1.49)

Note: Standard deviations are shown in brackets

Figure 11.3.1.1 **Observer ratings of the Accuracy of Good and Poor Witness for each Questioning Style**

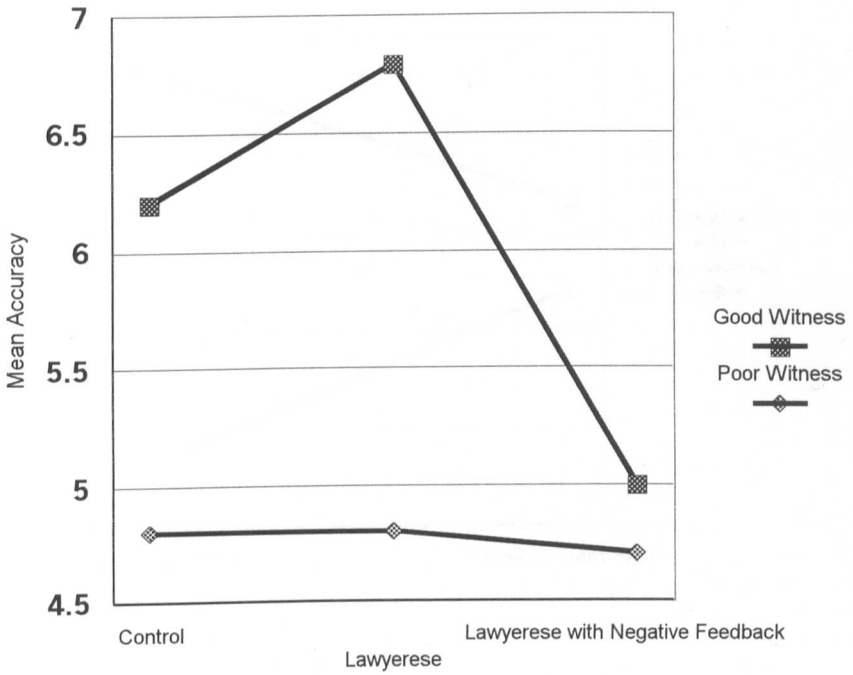
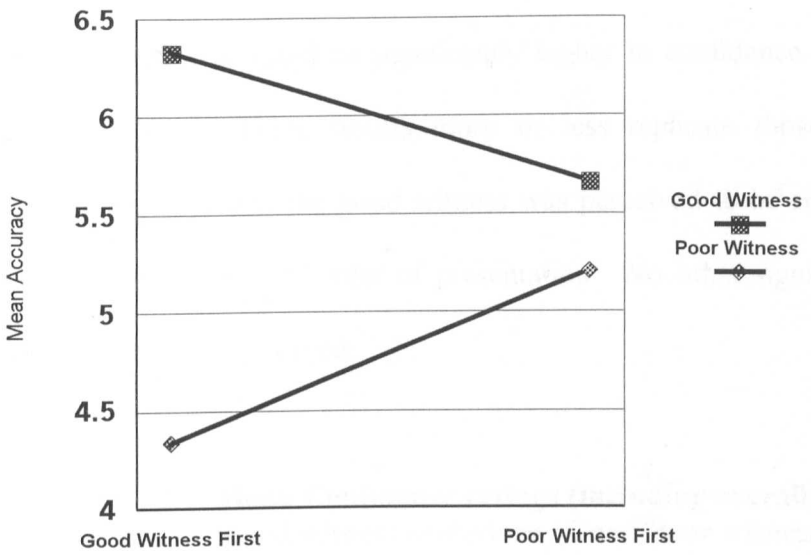


Table 11.3.1.2 Mean Accuracy ratings for taped witness interviews (Good/Poor witness) according to Order

ORDER	WITNESS	
	Good	**Poor**
Good First N=30	**6.33** (1.52)	**4.33** (1.21)
Poor First N=30	**5.67** (1.84)	**5.20** (1.32)

Note: Standard deviations are shown in brackets

Figure 11.3.1.2 **Observer ratings of Accuracy of Good and Poor Witness according to Order**



11.3.2 Confidence

The observers' confidence ratings for the two witnesses were also analysed by means of a 3 x 2 x 2 mixed ANOVA (Control; Lawyerese; Lawyerese with Negative Feedback X good witness first; poor witness first X good witness; poor witness) with repeated measures on the third factor. The means and standard deviations are shown in Tables 11.3.2.1 and 11.3.2.2. No main effect for confidence was observed between the three interview conditions, however a significant main effect was observed for order $F(1,54) = 5.16$, $p = .05$); confidence ratings were higher when the poor witness was rated first, see Table 11.3.2.2. A significant main effect was also observed for good/poor witness, $F(1,57) = 52.70$, $p = .001$; confidence ratings were higher for the good witness. The interaction (Order x good/poor witness) was again significant, $F(1,54) = 7.23$, $p = .01$, see Figure 11.3.2.2. F tests showed that,

within subjects, the good witness was rated significantly higher in confidence than the poor witness regardless of whether the good witness was presented first or second ($p<.01$). However, between subjects F-tests showed that the poor witness was rated as significantly higher in confidence when presented first ($p<.001$). These results more or less replicate those for perceived accuracy except that the good witness was perceived to be significantly more confident regardless of order of presentation. No other significant effects or interactions were observed.

Table 11.3.2.1 Mean Confidence ratings (including overall means) for taped witness interviews (Good/Poor witness) by questioning style

STYLE	WITNESS	
Overall 6.03 (1.16)	**Good** **6.93** (1.47)	**Poor** **5.13** (1.60)
Control N=20 **6.25** (1.25)	**6.90** (1.65)	**5.60** (1.73)
Lawyerese N=20 **5.90** (.85)	**6.80** (1.10)	**5.00** (1.45)
Lawyerese with Negative Feedback N=20 **5.95** (1.36)	**7.10** (1.65)	**4.80** (1.58)

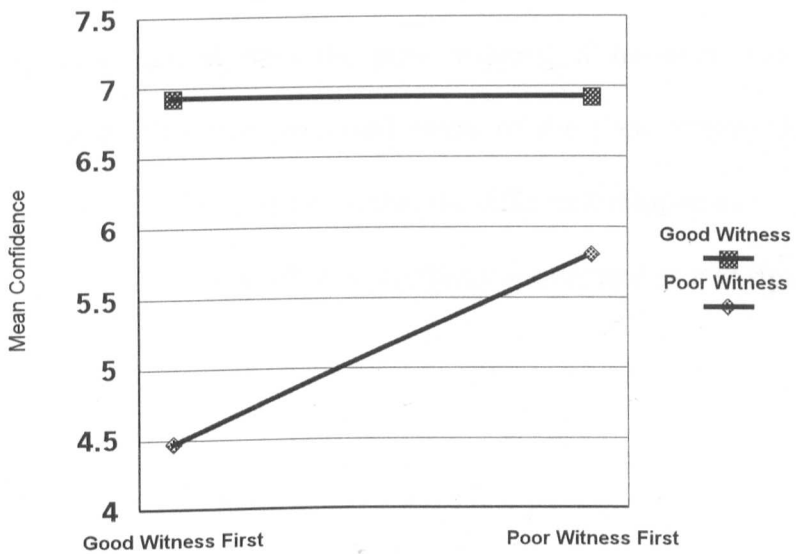
Note: Standard deviations are shown in brackets

Table 11.3.2.2 Mean Confidence ratings for taped witness interviews (Good/Poor witness) according to Order

ORDER	WITNESS	
	Good	**Poor**
Good First N=30	**6.93** (1.34)	**4.47** (1.57)
Poor First N=30	**6.93** (1.62)	**5.80** (1.35)

Note: Standard deviations are shown in brackets

Figure 11.3.2.2 Observer ratings of Confidence of Good and Poor Witness according to Order



11.3.3 Stress

The observers' stress ratings for the two witnesses were also analysed by means of a 3 x 2 x 2 mixed ANOVA (Control; Lawyerese; Lawyerese with Negative Feedback X good witness first; poor witness first X good witness;

poor witness) with repeated measures on the third factor. The means and standard deviations are shown in Tables 11.3.3.1 and 11.3.3.2. No significant main effects for stress were observed, although a trend was shown for good/poor witness, $F(1,54) = 3.79$, $p = .057$; highest stress ratings were for the poor witness (see Table 11.3.3.1).

The interaction (Order x good/poor witness) however, was significant, $F(1,54) = 6.49$, $p = .02$, see Table 11.3.3.2 and Figure 11.3.3.2. F-tests showed that within subjects, the good witness was rated significantly lower in stress than the poor witness, when the good witness was presented first ($p<.01$). However, no difference was observed between the good and poor witnesses' stress ratings when the poor witness was presented first ($p>.10$). None of the between subjects comparisons was significant. The means suggest that, if the good witness is presented first, the good witness is rated to be less stressed than the poor witness; if however, the poor witness is presented first, the perceived stress of the poor witness decreases, and the good witness increases such that the difference disappears.

None of the other main effects or interactions was significant.

Table 11.3.3.1 Mean Stress ratings (including overall means) for taped witness interviews (Good/Poor witness) by questioning style

STYLE	WITNESS	
Overall 2.85 (1.38)	**Good** **2.63** (1.67)	**Poor** **3.07** (1.65)
Control N=20 **2.60** (1.39)	**2.60** (1.67)	**2.60** (1.54)
Lawyerese N=20 **2.90** (1.41)	**2.40** (1.60)	**3.40** (1.90)
Lawyerese with Negative Feedback N=20 **3.05** (1.36)	**2.90** (1.77)	**3.20** (1.44)

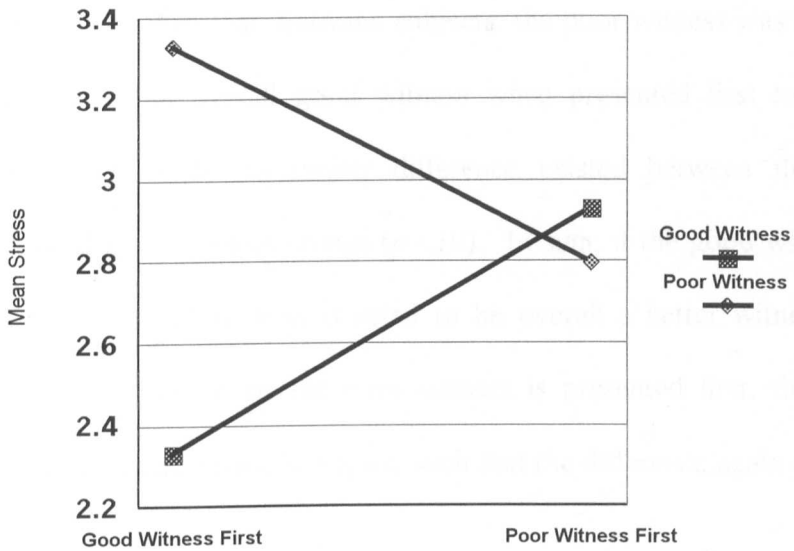
Note: Standard deviations are shown in brackets

Table 11.3.3.2 Mean Stress ratings for taped witness interviews (Good/Poor witness) according to Order

ORDER	WITNESS	
	Good	**Poor**
Good First N=30	**2.33** (1.42)	**3.33** (1.58)
Poor First N=30	**2.93** (1.86)	**2.80** (1.69)

Note: Standard deviations are shown in brackets

Figure 11.3.3.2 Observer ratings of Stress of Good and Poor witness according to Order



11.3.4 Good witness

The observers' overall 'good witness' ratings for the two witnesses were also analysed by means of a 3 x 2 x 2 mixed ANOVA (Control; Lawyerese; Lawyerese with Negative Feedback X good witness first; poor witness first X good witness; poor witness) with repeated measures on the third factor. The means and standard deviations are shown in Tables 11.3.4.1 and 11.3.4.2. No significant main effects between subjects for overall good witness were observed. However, a significant main effect was observed within subjects for good/poor witness, $F(1.54) = 23.57$, $p = .001$; good witness ratings were highest for the good witness.

The interaction (Order x good/poor witness) was also significant, $F(1,54) = 20.62$, $p = .001$, see Figure 11.3.4.2. F-tests showed that within subjects the good witness was rated significantly higher as an overall good

witness than the poor witness, when the good witness was presented first ($p<.001$). However, no difference was observed between the good and poor witnesses' ratings when the poor witness was presented first ($p>.10$). Further F-tests showed that, between subjects, the poor witness was rated significantly higher as an overall good witness when presented first rather than second ($p<.001$), while no similar difference existed between the good witness' overall good witness ratings ($p>.10$). In sum, if the good witness is presented first, the good witness is rated to be overall a better witness than the poor witness; if however, the poor witness is presented first, the perceived poor witness' competence increases, such that the difference again disappears.

Table 11.3.4.1 Mean Good Witness ratings (including overall means) for taped witness interviews (Good/Poor witness) by questioning style

STYLE	WITNESS	
Overall 5.38 (1.35)	**Good** **5.90** (1.58)	**Poor** **4.87** (1.72)
Control N=20 **5.60** (1.19)	**6.10** (1.21)	**5.10** (1.65)
Lawyerese N=20 **5.65** (1.46)	**6.30** (1.63)	**5.00** (1.84)
Lawyerse with Negative Feedback N=20 **4.90** (1.33)	**5.30** (1.75)	**4.50** (1.70)

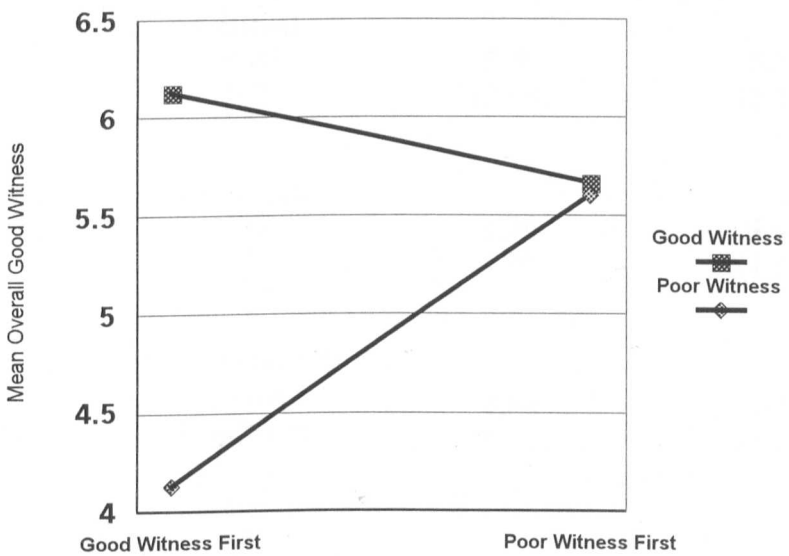
Note: Standard deviations are shown in brackets

Table 11.3.4.2 Mean Good Witness ratings for taped witness interviews (Good/Poor witness) according to Order

ORDER	WITNESS	
	Good	**Poor**
Good First N=30	**6.13** (1.46)	**4.13** (1.46)
Poor First N=30	**5.67** (1.69)	**5.60** (1.67)

Note: Standard deviations are shown in brackets

Figure 11.3.4.2 Observer ratings of Good Witness of Good and Poor witness according to 0rder



11.3.5 Fairness

The mean fairness ratings for each condition (i.e. Control; Lawyerese; Lawyerese with Negative Feedback) were then analysed using a 3 x 2 x 2

mixed ANOVA (Control; Lawyerese; Lawyerese with Negative Feedback X Order; good witness first; poor witness first X good witness; poor witness) with repeated measures on the last factor. No significant effects were shown in any of the analyses ($p>.10$). The means and standard deviations are shown in Tables 11.3.5.1 and 11.3.5.2. The means ranged from 4.00 to 5.30 indicating that all conditions were deemed to be of approximately average fairness.

Table 11.3.5.1 Mean Fairness of Questioning ratings (including overall means) for taped witness interviews (Good/Poor witness) by questioning style

STYLE	WITNESS	
Overall 4.62 (2.07)	**Good** **4.70** (2.11)	**Poor** **4.53** (2.13)
Control N=20 **5.25** (2.07)	**5.30** (2.08)	**5.20** (2.14)
Lawyerese N=20 **4.15** (1.84)	**4.30** (1.98)	**4.00** (1.89)
Lawyerese with Negative Feedback N=20 **4.45** (2.21)	**4.50** (2.24)	**4.40** (2.26)

Note: Standard deviations are shown in brackets

Table 11.3.5.2 Mean Fairness of Questioning ratings for taped witness interviews (Good/Poor witness) according to Order

ORDER	WITNESS	
	Good	**Poor**
Good First N=30	**4.33** (2.23)	**4.20** (2.13)
Poor First N=30	**5.07** (1.93)	**4.87** (2.09)

Note: Standard deviations are shown in brackets

11.4 DISCUSSION

In terms of jurors' perceptions of accuracy and confidence as a function of questioning style, the results of the fifth study indicate that observers seemed able to discriminate to some extent between the accurate and inaccurate witness accuracy in the Control and Lawyerese conditions but not in the Lawyerese with Negative Feedback condition. Moreover, because they really only had confidence judgments to go on, it seems reasonable to assume that the discriminations were based primarily on differences in perceived confidence. As mentioned previously, a variety of evidence suggests that confident witnesses tend to be judged as more accurate. These results suggest, therefore, that the subtle negative feedback style of questioning destroys the link that observers make between witness confidence and accuracy; the confident witness is not necessarily assumed to be more accurate.

However, the practical importance of this finding obviously rests crucially on the relationship between confidence and accuracy in any

particular situation. If the actual confidence-accuracy correlation is small or even negative in a particular situation it would obviously be spurious to judge accuracy from confidence. It is somewhat ironic therefore, that in experiment 2 the only condition to show an overall positive significant between subjects correlation between confidence and accuracy was the Lawyerese with Negative Feedback condition. In other words, in the one interviewing condition in which accuracy might reasonably have been inferred from confidence, the interviewing style negated such an inference by observers. On the other hand, although the general impression of the jurors' was that, overall, the witnesses in the negative feedback condition were the least accurate; this finding did not map exactly on to those for confidence, suggesting that jurors were not going entirely on perceptions of confidence when making their judgments of accuracy. It seems likely that they were simply very sceptical of this condition. Overall, therefore, it seems that, although observers may judge accuracy on the basis of confidence, accuracy judgments are further moderated by style of questioning, such that a confident witness in the negative feedback condition will not necessarily be assumed to be more accurate.

Moreover, successful inferences about relative confidence and accuracy between the good and poor witness disappeared when the poor/less confident witness was interviewed first; presumably because of difficulties in establishing an anchor or baseline for such judgments. When the poor/less confident witness came first, the perceptions of the witness' confidence and accuracy were inflated. Judgments of witness competence and stress were affected in much the same way; when the poor witness was presented first, perceptions of the poor witness' competence also increased, and perceptions of

stress decreased. With regard to perceptions of stress and witness competence, however, there were also trends, significant in the former, for the good witness to be affected in the opposite direction when the poor witness came first. There is no obvious explanation for this, though perhaps the failure to establish anchors on the previous judgments might have confused observers somewhat so they might have adopted a 'go for the mean' strategy in these later ratings.

No significant effects were found for fairness; this suggests, perhaps, that Lawyerese questioning styles in themselves (i.e. not coupled, for example, with aggressive intonation or nonverbal gestures) may possess sufficient subtlety not to be perceived as intimidatory and unfair.

CHAPTER 12

PREDICTING A GOOD WITNESS UNDER CROSS-EXAMINATION: Study 6

12.1 INTRODUCTION

Given that witnesses are likely to differ in their responses during an interrogative interview, questionnaire data were collected during the second and fourth experiments to examine possible predictors of witness credibility in court.

12.2 Data collected as part of Experiment 2

The rationale for the inclusion of the measures in experiment 2 was as follows.

Social distress

Many researchers have been interested in social anxiety as an interpersonal phenomenon (Alden, 1989; Borkovec, Robinson, Pruzinsky, & DePree, 1983; Britt, 1992; Pruzinsky & Borkovec, 1990). Virtually everyone experiences social anxiety occasionally. When they do, people suffer not only subjective tension but may behave in ways that might interfere with social interaction. Also, when nervous, people tend to display overt indications of their inner arousal (e.g. trembling, fidgeting), avoidance of other people, and disruption of other ongoing behaviors (e.g. irregular speech, difficult concentration). As a result, anxiety can be a liability in social relations, because people who are socially distressed may become less socially effective.

While most research has been directed toward understanding social anxiety and its impact on interpersonal behavior, other researchers have been interested in

social anxiety in the process of studying other phenomena. For example, the construct of social anxiety has been used in studies of topics such as evaluation apprehension, impression management, self-consciousness, affect regulation, self-efficacy, alcohol abuse, conformity, and loneliness. This research has demonstrated that feelings of social inadequacy and concerns about negative evaluations by others play a central role in many psychological phenomena (for reviews see Jones, Cheek, & Briggs, 1986, Leary, 1983b). Thus people's feelings about others' evaluations may have a role to play in interrogative interview situations.

Watson and Friend (1969) have defined fear of negative evaluation (FNE) as "apprehension about other's evaluations, distress over their negative evaluations, and the expectation that others would evaluate one-self negatively" (p. 449), and have constructed a scale that assesses the degree to which people worry about how they are perceived and evaluated by others. The FNE scale focuses primarily on people's concerns with interpersonal evaluation rather than on the tendency to feel anxious *per se*. However, to the extent that social anxiety is closely related to people's concerns with how they are perceived and evaluated by others, fear of negative evaluation and social anxiety are closely related constructs (Schlenker & Leary, 1982).

Two versions of the FNE scale exist. The original FNE scale (Watson & Friend, 1969) consists of 30 true-false items approximately balanced between positively and negatively scored items. The revised, brief version of the scale (Leary, 1983a) contains 12 of the original items, which are answered on five-point scales (1, not at all characteristic of me; 5, extremely characteristic of me). The opposite of high FNE is the lack of apprehension about others' evaluations, but not necessarily a desire or need to be evaluated positively.

In order to measure social avoidance and distress more specifically, Watson & Friend (1969) developed the Social Avoidance and Distress (SAD) scale. Watson and Friend propose distinct avoidance and distress subscales, and a factor analysis by Patterson and Strauss (1972) revealed separate behavioural and anxiety factors. Nevertheless, one study has shown that the avoidance and anxiety subscales differentially predict behaviour in real interactions (Leary, Knight, & Johnson, 1987).

In view of these considerations, one might predict that people who score high on the Fear of Negative Evaluation (FNE) will be prone to become apprehensive in courtroom examination situations and attempt to avoid social disapproval. Consequently they might more vulnerable to leading questions. Indeed Gudjonsson (1998; 1992) found that high FNE scores related to suggestibility. As a consequence, there might be detrimental effects on confidence and accuracy relationships.

In addition, subjective feelings of Distress about such an interview, and the Discomfort a witness may feel during this kind of interrogative situation might also influence cognitive performance; however, the effects might not necessarily be detrimental. For example, Nolan and Markham (1998) found that highly anxious participants showed greater positive confidence-accuracy correlations; they suggested that this occurred because highly anxious people were more likely to carefully appraise their performance.

To test these predictions, the FNE, the Distress subscale of the SAD and a Discomfort scale devised by the experimenter were used.

12.3 METHOD (Data collected as part of experiment 2)

12.3.1 Participants, materials and procedure

The participants, materials and procedure were identical to those used in experiment 2 (see Chapter 8).

As mentioned earlier, 60 participants were randomly assigned to one of three interview conditions (Control, Lawyerese, Lawyerese with Negative Feedback), and were required observe a colour video of 4-5 minutes duration depicting a criminal offence. Following the completion of a 5 minutes filler task participants were then asked a series of questions that required 'yes' or 'no' answers, according to the allocated questioning style. At the end of the interview schedule, all participants completed a series of questionnaires and self-reports.

The questionnaire included: a) the Distress subscale of the Social Avoidance and Distress (SAD) Scale (14 items); the Guttman split-half reliability was .82 on the present sample; and, b) the 'Brief' Fear of Negative Evaluation (FNE) Scale (12 items); split-half reliability = .82. The 26 items were randomly combined to make up one questionnaire.

A short six-item Discomfort scale was also included. Participants were asked to decide the degree to which statements were characteristic of them in the situation to which they were assigned on a Likert scale where 1 represented 'not characteristic of me', and 5 represented 'extremely characteristic of me'. Typical statements were 'I did not feel tense or nervous about answering the questions'; 'I was relieved when the interview had finished'. The Guttman split-half reliability of this scale on the present sample was .66. The scale was designed to measure subjective stress at retrieval (see also Appendix 7).

To gather some more general information, participants were also asked to rate themselves on a Likert scale ranging from 1 to 5 on the following measures; a) how accurate they thought they were; where 1 represented 'extremely inaccurate' and 5 represented 'extremely accurate'; b) how confident they were in their performance; where 1 represented 'not at all confident' and 5 represented 'extremely confident'; c) how good a witness they thought they were; where 1 represented 'extremely poor' and 5 represented 'extremely good'; d) how happy they would be to give evidence in court; where 1 represented 'not at all happy' and 5 represented 'extremely happy'; e) how well they thought they might cope with giving evidence in court; where 1 represented 'not at all well' and 5 represented 'extremely well'; and, f) how good a witness they thought they would be in court; where 1 represented 'extremely poor' and 5 represented 'extremely good'; see Appendix 7).

At the end of the procedure, all participants were thanked for their participation and debriefed fully as to the nature of the research, and given the opportunity to ask questions or raise issues for discussion.

12.4 RESULTS

One-way ANOVAs (Control, Lawyerese, Lawyerese with Negative Feedback) were performed separately on the data for distress (mean scores on Likert scale 1-5), fear of negative evaluation (mean scores on Likert scale 1-5), and discomfort (mean scores on Likert scale 1-5). No significant effects were observed in any of the analyses ($p>.10$). The mean total scores and standard deviations are shown in Table 12.3.1.

Table 12.4.1 Mean Distress, Mean Fear of Negative Evaluation (FNE) and Mean Discomfort by Questioning Style

STYLE	Distress	FNE	Discomfort
Control N=20	**34.40** (10.76)	**37.25** (6.52)	**15.65** (4.07)
Lawyerese N=20	**35.75** (9.12)	**37.35** (9.65)	**16.65** (3.36)
Lawyerese with Negative Feedback N=20	**32.55** (9.72)	**32.85** (8.33)	**15.35** (3.23)

Note: Standard deviations are shown in brackets

As the main purpose of this study was exploratory, Pearson's correlations were performed between the following variables: discomfort; fear of negative evaluation; social distress; confidence-accuracy correlation for overall interview; confidence-accuracy correlation for easy, moderate and difficult items, separately; number correct for overall interview, easy, moderate and difficult items, separately; mean confidence for overall interview, easy, moderate and difficult items, separately; hits for overall interview and difficult items, separately; false-alarms for overall interview and difficult items, separately; percentage accuracy for overall interview and difficult items, separately; self-reports relating to accuracy, confidence, good-witness, evidence-giving in court, coping in court, and good-witness in court (as outlined in 12.2.1 above). Because of the large number of correlations a fairly conservative significance level was adopted ($p < .01$).

No significant relationships between the questionnaire measures and measures of accuracy, confidence and C-A relationships were found. The only exception was that those who considered themselves most accurate were also more likely to produce more correct responses overall (r=.36, n=60 p<.01); see Appendices 12 and 13.

12.5 Data collected as part of Experiment 4

Given that absence of significant results using the measures in experiment 2 most of the individual difference measures were dropped for experiment 4, and the emphasis was changed. The rationale for the measures used in experiment 4 was as follows.

Verbal fluency

Another possible source of individual differences in cognitive performance in the courtroom is neuropsychological function. Good confidence and accuracy relationships require careful monitoring and calibration of memory; a job that would seem to imply activation of the frontal lobes (Troyer, 2000; Troyer, Moscovitch & Winocur, 1997; Troyer, Moscovitch, Winocur, Alexander & Stuss, 1998). Tests of verbal fluency (i.e. a measure of the ability to successfully retrieve, recognise, and/or correctly combine words verbally, as opposed to in writing) are frequently used in clinical and experimental examinations of neuropsychological and cognitive function.

While the most commonly used score from verbal fluency tests is the total number of words generated, this gives little information about the underlying cognitive processes involved in fluency performance. To this end, separate behavioural components have been operationalised as 'switching' and 'clustering'

(Troyer, Moscovitch & Winocur, 1997). Clustering is thought to be a relatively automatic process, related to temporal lobe functioning. Switching, however, is related to frontal functioning and involves cognitive flexibility in shifting from one subcategory to another and is a relatively effortful process.

If good confidence-accuracy relationships require substantial frontal activation, one might predict that high scorers on the switching component of verbal fluency might show the highest accuracy, and confidence-accuracy correlations.

12.6 METHOD (Data collected from experiment 4)

12.6.1 Participants, materials and procedure

The participants, materials and procedure were identical to those used in experiment 4 (see Chapter 8).

Firstly, the participants were required to undergo a neurological test of verbal fluency. The test of verbal fluency was a phonemic task taken from Troyer (2000). Each individual protocol was tape recorded for later transcription. For the phonemic fluency test, participants were instructed to generate out loud words beginning with *f*, *a*, and *s*, excluding proper names and variants of the same word (e.g., the same word with different suffixes). Sixty seconds was allotted for each of the three phonemic trials.

Three scores were calculated on each phonemic fluency test: (a) number of words generated, excluding errors and repetitions, (b) mean cluster size, and (c) number of switches.

Detailed rules for scoring switching and clustering are provided in the Appendix (see Appendix 14). Briefly, on phonemic fluency, clusters were defined as groups of successively generated words that began with the same first two letters (e.g. *art*, *arm*), differed only by a vowel sound (e.g. *sat*, *seat*, *soot*), rhymed (e.g. *sand*, *stand*), or were homonyms (e.g. some, sum), when indicated by the participant as being differently generated words.

Cluster size was counted beginning with the second word in each cluster, and the mean cluster size was also calculated for the phonemic test. Switches were calculated as the number of transitions between clusters, including single words, for the phonemic fluency test.

Following transcription, each protocol was then scored by the interviewer, and half of the protocols were scored for cluster size and number of switches by an independent rater. Inter-rater reliabilities, calculated with Pearson coefficients, were high for phonemic fluency cluster size, $r(10) = .99$, and switching, $r(10) = .99$.

It should be noted that, in this study, the verbal fluency test was administered prior to the first interview (i.e. it was not administered again prior to the interview after 6 months).

The materials and procedure used following this were identical to those previously described for the standard Lawyerese condition (see Chapter 8 and Chapter 10). To reiterate, all participants were told that they were to observe a videotape of an event sequence of 4-5 minutes duration, and afterwards they would be required to answer some questions. All participants were then shown the 5 minutes colour video depicting a criminal offence. Following this every participant was required to complete a filler task for a period of five minutes, which involved reading unrelated material. Participants were advised that they would now be asked

a series of questions regarding the videotape they had observed. All questions required 'yes' or 'no' answers, and following each question participants were asked to rate their confidence in the response they had given on a Likert scale ranging from 1 to 9, where (1) represented 'pure guess' and (9) represented 'absolutely certain'.

Each participant was then individually interviewed under the standard Lawyerese condition. As mentioned previously, in the Lawyerese condition participants were required to give answers to the three sets of target items (i.e. 14 'easy', 14 'moderate, and 10 'difficult' items), but the questions were so phrased as to replicate the manner in which lawyers conduct cross-examinations in court.

At the end of the interview schedule, at both 5 minutes and 6 months retention intervals all participants completed the discomfort questionnaire and self-reports as described in 12.2.1 (see Appendix 7).

12.7 RESULTS

The descriptive data for the phonemic fluency task were as follows. The mean cluster size was .35 (SD = .17); mean switching score was 27.43 (SD = 9.68); and the mean total word score was 39.81 (SD = 12.09).

Pearson's correlations were performed between the following variables: mean total word score on FAS; mean number of switches on FAS; mean cluster score on FAS; discomfort at 5 minutes and 6 months, separately; accuracy for overall interview at 5 minutes and 6 months, separately; accuracy for easy items at 5 minutes and 6 months, separately; accuracy for moderate items at 5 minutes and 6 months, separately; accuracy for difficult items at 5 minutes and 6 months, separately; confidence for overall interview at 5 minutes and 6 months, separately;

confidence for easy items at 5 minutes and 6 months, separately; confidence for moderate items at 5 minutes and 6 months, separately; confidence for difficult items at 5 minutes and 6 months, separately; confidence-accuracy correlation for overall interview at 5 minutes and 6 months, separately; confidence-accuracy correlation for easy items at 5 minutes and 6 months, separately; confidence-accuracy correlation for moderate items at 5 minutes and 6 months, separately; confidence-accuracy correlation for difficult items at 5 minutes and 6 months, separately; confidence in correct answers for overall interview at 5 minutes and 6 months, separately; confidence in correct answers for easy items at 5 minutes and 6 months, separately; confidence in correct answers for moderate items at 5 minutes and 6 months, separately; confidence in correct answers for difficult items at 5 minutes and 6 months, separately; confidence in incorrect answers for overall interview at 5 minutes and 6 months, separately; confidence in incorrect answers for easy items at 5 minutes and 6 months, separately; confidence in incorrect answers for moderate items at 5 minutes and 6 months, separately; confidence in incorrect answers for difficult items at 5 minutes and 6 months, separately; self-reports relating to accuracy, confidence, good-witness, evidence-giving in court, coping in court, and good-witness in court (as outlined in 12.2.1 above) at 5 minutes and 6 months each, separately.

Again, because of the large number of correlations a fairly conservative significance level was adopted ($p < .01$).

With regard to cognitive performance, again, few significant relationships were found. The exceptions were as follows.

Perhaps most important, as hypothesised, high switching scores (reflecting frontal activation) were significantly related to positive confidence accuracy

relationships for difficult items at 5 minutes (.41, n=21, $p<.01$, one-tailed), and accuracy overall at 6 months (.48, n=21, $p<.01$, one-tailed). Though high switching scores were also significantly related to confidence overall in difficult items at 6 months (.62, n=21, $p<.01$), confidence overall at six months (.52, n=21, $p<.01$) and confidence in incorrect answers overall (5 minutes and 6 months combined (.58, n=21, $p<.01$). See Appendices 15, 16, 17, 18, and 19).

12.8 DISCUSSION

Unfortunately little of interest came from this investigation of individual differences with the possible exception of the findings relating to verbal fluency. Potentially it seems that verbal fluency switching and correspondingly, a capacity for frontal lobe activation, may relate to increased positive confidence-accuracy relationships in the short-term, and greater accuracy (in terms of correct answers) in the long-term. In the long-term, however, it seems that, although the accuracy of those high on switching fluency remains high relative to those low on fluency, the C-A trend is modified such that confidence remains high, regardless of accuracy. These findings may simply reflect a statistical artefact because of the large number of analyses involved, however, they might alternatively indicate perhaps the interaction of two behavioural factors in those high on switching ability; i.e. an ability to remember accurately, calibrate accuracy with confidence, and a general confidence in their own ability. In the short-term, the ability to calibrate accurately might be sufficient to take precedence over the tendency to be overconfident. Over time, however, the overconfidence becomes more prominent.

Although much is known about the behaviour of patients with frontal lobe damage who score poorly on verbal fluency (Troyer *et al.*, 1997; 1998), little is

known about the personality and social characteristics of those scoring high on verbal fluency within the normal population. Perhaps the data here point to some interesting directions for future research. In terms of predicting courtroom performance, however, the present results suggest that possessing high verbal fluency ability may be a mixed blessing. At short intervals it may be useful, aiding confidence-accuracy relationships, after an interval, however, the result may be overconfidence even though those high on verbal fluency produce more correct answers than those low on fluency.

It should be argued that the method would have been more sensitive had the verbal fluency test been given again prior to the second interview. In retrospect this would probably been advisable; in fact, surprisingly, no reliability data are available for phonemic fluency; the assumption in clinical work seems to be that that frontal lobe capacity is fairly invariant in the short-term. Nevertheless, this possibility awaits further investigation.

Results relating to the effects of stress and fear of evaluation were clearly non significant; arguably these indicators might have been significant in a real life courtroom context, but there were no obvious trends in the data.

CHAPTER 13

REVIEW AND GENERAL DISCUSSION

At the beginning of this thesis it was noted that some have expressed concern that court officials lend too great a weight to evidence given by witnesses when reaching decisions, believing them to be accurate and thus truthful. Moreover, one of the most influential measures used by courtroom officials to determine accuracy is the confidence expressed by witnesses in respect of their testimony. Nevertheless, despite the large psychological literature on eyewitness accuracy and confidence, little empirical research has been conducted on these issues as they relate to performance in the courtroom, in particular, during cross-examination. Also, although much academic and legal debate has centred upon cross-examination, the focus has tended to be on witnesses who are classed as 'vulnerable', leading to a virtual vacuum of empirical investigation into cross-examination and its effects upon ordinary witnesses, who are statistically much more likely to be called upon to give evidence in court.

In view of these considerations, the central focus of this thesis was to investigate how ordinary witnesses perform under experimental cross-examination conditions. To reiterate, four main questions were investigated: 1) What are the effects of cross-examination questioning styles, including the provision of negative feedback, on witness confidence and accuracy, and the relationship between C-A? 2) What influence does time have on the stability

of confidence and accuracy, and C-A relationships, derived under cross-examination conditions? 3) How do cross-examination questioning styles influence jurors' assessments of witness confidence and accuracy? And 4) Can we predict who will be a good witness under cross-examination conditions?

13.1 The effects of cross-examination questioning styles on witness accuracy and confidence

As a starting point for the programme, on the basis of the psychological literature of eyewitness testimony, it was predicted that certain types of question (leading and declarative questions); and questioning style (lawyerese and lawyerese with negative feedback) would tend to reduce witness accuracy, confidence, and the C-A relationship. In Experiment 1, however, and contrary to some evidence suggesting that questioning types and styles used during cross-examination might impede witness accuracy (Dent, 1978; Dent & Stephenson, 1979; Kebbell & Giles, 2000; Lipton, 1975; Loftus, 1975; Marquis, Marshall & Oskamp, 1972; Turtle & Wells, 1988; Westcott, 1995), no negative effects on accuracy, confidence, or C-A, of question type or questioning style were found. Indeed, what non-significant trends there were, indicated that negative feedback tended to increase accuracy and the C-A relationship.

These results apparently conflict with Schooler and Loftus' (1986) theory of discrepancy detection in respect of negative feedback. This theory suggests that negative feedback will reduce confidence in memories making it less likely that suggestions made by a questioner will be detected, leading to less accurate responses. In Experiment 1, however, neither confidence nor

accuracy were reduced in response to negative feedback, in fact, the highest number of highly confident/correct answers was shown in the negative feedback condition. This also fits with other research showing that, in general, highly confident responses tend to be accurate (Kebbell, Wagstaff & Covey, 1996).

However, it is of interest that Schooler and Loftus' theory of discrepancy detection does not differentiate between the types of feedback that might be employed in interrogative interviewing situations; for example, overt and subtle. As mentioned in Chapter 3, an alternative prediction might be derived from Gudjonsson & Clark's (1986) model of interrogative suggestibility. From this perspective, negative feedback, if negatively appraised, will be rejected as suspicious. In Experiment 1 the feedback was overt or transparent; it is possible, therefore, that participants found this feedback suspicious and rejected it; this rejection, in turn, might have reinforced their confidence in items for which they already had a degree of certainty; i.e. the process of rejecting attempts to mislead them might have exaggerated their confidence in items for which they might otherwise have expressed slightly less certainty; hence they showed the highest number of highly confident/correct answers.

Suspicion might also have arisen from the random allocation of question items to questioning style conditions, together with the requirement for equal numbers of 'yes' and 'no' answers in each questioning condition; in many cases this would almost certainly have signalled the purpose of the interview to the witness.

To overcome these constraints on ecological validity, Experiment 2

was conducted. This interview was specifically designed and constructed so as to replicate as far as possible the way in which counsellors might conduct cross-examinations. Accordingly, three separate conditions were again adopted, control, lawyerese, and lawyerese with negative feedback; but this time for the lawyerese condition, phrasings for the questions were taken directly from Crown Court cross-examination transcripts and allowed for the questions used to replicate the manner in which lawyers conduct cross-examinations in court. This kind of questioning prompts desired answers and required more affirmative responses. The lawyerese with negative feedback condition employed an identical interview schedule to that used for the lawyerese condition except that subtle negative feedback was applied to each 'no' answer, regardless of accuracy, to reassert an expectancy of 'yes' responses. In addition, item difficulty was also systematically investigated. Chapter 5 indicated that one of the crucial variables that could affect witness accuracy and confidence was item difficulty (Kebbell, Wagstaff & Covey, 1996); if item variance is increased through the inclusion of items that are easier to remember (sex, whether it was day or night), confidence-accuracy relationships tend to improve.

Additionally, and importantly, in order to further enhance and maintain ecological validity the schedule was specifically constructed with an imbalance between the numbers of items in the item difficulty categories, and the requirement for 'yes' and 'no' responses. Given this methodological asymmetry, it may be useful to go through the rationale again.

An examination of courtroom transcripts indicated that practising lawyers tend to phrase many of the questions during cross-examination in such

a way that 'yes' responses are actively encouraged. This then makes it far more difficult for witnesses to switch to a 'no' response on the fewer critical items, upon which, in the main, the cross-examining lawyers' case may ultimately rest. Moreover, generally speaking, such critical items are those that have been identified as inconsistent, both within and between witnesses' statements. Consequently, one finds that, in practice, the critical items are those items about which there may be some ambiguity (i.e. such ambiguity represents a perfect opportunity for examiners to discredit the testimony of the witness during cross examination), and this is most likely to be the case for items that are actually difficult to remember.

Also, from the cross-examining lawyers' point of view, the strategy of targeting difficult items makes perfect sense, as cross-examiners would be considerably less likely to influence the witness' response to a very easy item, about which they would most probably be very confident in. In addition, it is apparent from the transcripts that lawyers will tend to target only a few critical items, in turn, preventing their strategy from becoming too transparent. Thus, when wishing to cast doubt on the reliability of a particular witnesses' testimony overall, they often target a few difficult items, setting these up as exemplars in the interviewing framework described. This procedure was therefore adopted in Experiment 2.

Experiment 2 found that participants were only significantly less accurate in their responses when the critical items were classed as 'difficult' and when subtle feedback was present during the lawyerese interview. As noted previously, one possible interpretation is that subtle (but not necessarily overt) negative feedback, when applied to difficult items, reduces discrepancy

detection; i.e. because the items are those about which the witness is uncertain and the subtle feedback enhances this uncertainty.

However, the effects of the questioning style and item difficulty variables on C-A correlations are rather more complex to explain. In relation to the overall data, and that for easy and moderate items, no detrimental effects of lawyerese questioning were found, regardless of whether feedback was applied. In fact, some significant and positive between subjects correlations were found only for these conditions. On the other hand, for difficult items, when the standard lawyerese condition alone was used, participants' C-A relationships actually became negative (though not significantly so) in direction, due primarily to a significant reduction in confidence in correct answers. However, contrary to predictions, no similar reduction in the C-A relationship was shown when the subtle feedback was added. Indeed, the means for confidence in correct and confidence in incorrect answers indicate that the significant negative effects of the lawyerese condition on confidence in correct responses were cancelled by the challenge of subtle feedback. .

The explanation given in Chapter 8 was as follows. The lawyerese questioning style first reduced confidence in correct answers. However, when the witness said 'no' to any question, subtle feedback was applied to attempt to change the response to 'yes'. This might have given the impression to the witness that whenever he or she gave 'no' as an answer a challenge would follow. Consequently, a witness might only say 'no' when he or she was relatively certain he or she was correct, effectively cancelling out the initial negative effect to some degree.

Broadly speaking, therefore, the results suggest that witnesses may be

more strict in applying a high confidence judgment when subtle feedback is present. This could account for the significant positive between subjects C-A relationships found for the overall combined data and for the easy items (the former was clearly due to the latter) in this condition. That is to say, when the items were easy, participants maintained their confidence in correct answers despite the negative feedback (because the answers were obvious) but decreased their confidence in those answers that were incorrect (less obvious), the latter being more amenable to influence by negative feedback.

It should be noted that an apparently aberrant result occurred in the lawyerese (without feedback) condition for moderate items; this was the only condition to show a significant between subjects C-A correlation. The reason for this is unclear; it may simply be an artefact, but the finding illustrates well the fact that the lawyerese conditions appear to have no negative effects on easy and moderate items.

Taken together, the results suggest that standard types of interview procedure employed in cross-examination produce mixed effects in terms of accuracy and C-A calibration. Thus the lawyerese style alone has little effect on accuracy generally, but a negative effect on calibration (C-A) for critical items, whereas addition of subtle negative feedback to some extent overcomes the C-A calibration problem, but at the expense of accuracy for critical items.

As mentioned above, Experiment 2 was asymmetrical in its design in order to enhance its ecologically validity. One aspect of this was that correct responses to 'difficult' items required a 'no' response. To investigate whether the significant findings in Experiment 2 occurred solely because a 'no' response was required, Experiment 3 was conducted, which substituted the 10

difficult items used in Experiment 2 with 10 additional easy questions that required a 'no' answer to be correct, to test the effects of 'nay-saying' *per se*. If the effects were solely due to saying 'no', or influenced by this factor, they should occur primarily for the easy items requiring a 'no' answer.

Experiment 3 found no such effects. On the whole, these data suggest the significant effects found for the 'difficult' items occurred primarily because of the nature of the material in the question rather than because a 'no' response was necessary for a correct answer. Of additional interest, however, were the impressive between subjects correlations for the control and lawyerese with negative feedback conditions shown also in the analyses of confidence in correct and in incorrect answers. Once again, this supports recent findings that confidence-accuracy relationships may not necessarily be as poor as early researchers have suggested (Kebbell, Wagstaff & Covey, 1996). Moreover, the fact that correlations of a similar magnitude were not found in Experiment 2 would seem to have occurred because of the deleterious effects of difficult items in Experiment 2. On this note, C-A discrepancies in the literature are most probably due to item difficulty and thus reflect rather more the materials used. In contrast, a substantial but non-significant negative between subjects confidence-accuracy correlation for easy items for the lawyerese condition (see Table 9.3.2) was observed which might be worthy of further investigation. The reason for this is again, unclear, but given that no similar effect was found for overall items (moderate items included), this could have been an artefact.

If valid and generalisable, the results of these three experiments overall may have implications for theory, research and practice. For example, the

need to differentiate between types of negative feedback (i.e. overt and subtle) is important and is not addressed by current theoretical approaches in this area. As noted, Gudjonsson and Clark's (1986) model of interrogative suggestibility and Schooler and Loftus' (1986) theory of discrepancy detection have not specified the types of negative feedback that are available to interrogative interviewers and their differential effects. Future research should include this differentiation.

On a practical level, it has been noted that some legal commentators have strongly defended and advocate the use of strategies such as leading questions and feedback during cross-examination (Stone, 1988). However, in general, the results of Experiment 2 indicate that the rationale upon which the cross-examination procedure rests can now be seriously questioned. If cross-examiners are really interested in eliciting accurate responses from witnesses in response to critical items it seems that lawyerese questioning alone does not help, and the use of a subtle feedback strategy in addition to a lawyerese questioning style is certainly not the way to achieve it.

Having said this, however, it is essential to emphasise the limitations of the present studies in terms of the characteristics of the sample. For example, it is important to assess the effects of overt and subtle feedback upon those witnesses who would be classed as vulnerable to this kind of interviewing (i.e. child; learning disabled; older adults). Perhaps even more important, however, the participants in these experiments were presumably willing to do the best they could in terms of producing accurate testimony and rating their confidence; i.e. they did not set out to deliberately deceive. In real courtroom situations, however, lawyers are not simply concerned with retrieving the

maximum accurate information from willing witnesses, but also 'catching out' those who may deliberately be attempting to hide or distort information. The present studies did not address the effectiveness of the lawyerese styles in obtaining information from unwilling and untruthful witnesses. Clearly future research should investigate this issue.

13.2 The effects of delay on the stability of confidence and accuracy, and C-A relationships, derived under cross-examination conditions?

Given the recent Home Office emphasis on reducing delays in the Criminal Justice System (Home Office, 1999, Glidewell Recommendations 21 and 22 in the Magistrates' Courts), including the time between initial police questioning and subsequent appearances in court, and also the time between initial appearances in court and appeals, research on the effects of delay on eyewitnesses' accuracy, confidence, and the relationship between confidence and accuracy was also investigated in present thesis.

As noted in Chapter 11, only a few studies have examined the effects of delay. Two studies have shown that repeat questioning of the same material can increase confidence after 2 days (Hastie, Landsman & Loftus, 1978), and after 3 weeks (Turtle & Yuille, 1994). Another, however, reported that confidence decreased following repeat questioning after 1 week, and more so for incorrect than correct answers, resulting in an increase in confidence-accuracy judgements over time (Ryan & Geiselman, 1991). More recently, a study carried out by Granhag (1997) suggests that the positive effect of delay (1 week) on confidence-accuracy may be more a function of making repeated confidence ratings *per se*, than the delay involved. Thus far then, the results

on repeated recall on confidence are contradictory, and no work has examined these effects over a time period more akin to that experienced by witnesses in the Criminal Justice System, or examined the effects of item difficulty. Experiment 4 therefore aimed to illuminate the literature with regard to these issues using only a standard lawyerese condition (without negative feedback). This condition was chosen a) it most closely reflects cross-examination as might be conducted in the courtroom, and b) the lawyerese interview alone had the most detrimental effect on confidence-accuracy relationships. In Experiment 4, therefore, a new sample of participants was presented with the same lawyerese interview materials as used in Experiment 2, and then tested after 5 minutes and 6 months. As new participants were used in Experiment 4 their results were also compared with the results of Experiment 2 as a reliability check.

It was hypothesised that, given standard theories of memory decay, both confidence and accuracy would reduce over a long delay. Not surprisingly, the results of Experiment 4 showed that a 6 months delay significantly reduced accuracy for all levels of question difficulty. Similarly, this amount of delay also reduced confidence for all levels of question difficulty, except the easy items. Confidence and accuracy were also highest for easy compared to moderate items; however, these trends were not differentially affected by the delay.

However, importantly, and as previously pointed out in Chapter 10, it is also feasible that memory might decay selectively; for example, data from this thesis and previous studies (Kebbell, Wagstaff, & Covey, 1996) indicate that when participants are absolutely sure of a response they tend to be

accurate. Arguably, such responses are the least likely to decay over time, not only because of the strength of the original association, but also because they might be the most obvious candidates for rehearsal. In contrast, other responses that have weaker associations might be more likely to decay over time, with a corresponding decline in confidence. One result of this might be a decline in the overall number of correct responses over time, but an increase in confidence-accuracy relationships as the distinction between the two types of response (accurate sure, inaccurate unsure) becomes more exaggerated. This would account for Ryan and Geiselman's (1991) result that confidence-accuracy increased after a delay largely due to the lowered confidence in incorrect answers.

The within subjects C-A relationships were not significant; nevertheless, analyses of confidence in correct and incorrect answers showed, overall, a significant tendency for greater confidence in correct answers for the easy and moderate items, and in incorrect answers for the difficult items. However, no differential effects of delay were found. Thus Experiment 4 showed no support for Ryan and Geiselman's (1991) finding that a delay led to increased within subjects C-A relationships due to lowered confidence in incorrect answers, or indeed, Granhag's (1997) finding that within subjects C-A correlations increased after a delay for participants who had been given an opportunity to review their previous confidence ratings (i.e. had been shown their original ratings).

A formal review of previous confidence ratings was considered inappropriate in Experiment 4 because of considerations of ecological validity. It is quite possible, however, that under Granhag's formal review condition

participants tended to stick to their previous confident responses, but were more willing to reconsider inaccurate responses about which they were much less confident, thus leading to lowered confidence in incorrect answers. Arguably, therefore, if a formal review of confidence had been performed at 6 months similar results might have been found in the present study.

Of course there are occasionally situations in the courtroom in which witnesses do have some opportunity to review what they have previously said, such as when a police officer is allowed to consult his or her notebook; however, this is not in any sense equivalent to a formal review of confidence judgments. Taken together, therefore, the present results and those of Granhag suggest that, normally, within subjects C-A relationships are not affected by a delay, however, perhaps they could be improved if specific measures of confidence were taken at the time of the initial interview, and witnesses were allowed to consult them in court. The practical problems, however, in putting such measures into practice, would, however, seem considerable. Moreover, the applicability of Granhag's findings would seem to rest crucially on the validity of the witnesses' original C-A assessments. If their original within subjects C-A judgments were poor or even negative, the effects of confirming their confident responses would result even greater calibration inaccuracy.

While within subjects C-A correlations in Experiment 4 remained non-significant and unaffected by the delay, between subjects correlations became significant and positive for the overall data (question difficulty levels combined) and for the easy and moderate questions separately. In contrast, the between subjects C-A correlation for difficult questions after the delay was slightly more negative, though not significantly so. The interpretation offered

in Chapter 10 for these trends was as follows. The trend for the between subjects correlations to be exaggerated over time was not a consequence of lowered confidence in incorrect answers relative to correct answers. Instead, a closer examination of the data showed evidence of greater sample variability for both accuracy and confidence score for all levels of item difficulty after the delay. The increases in heterogeneity would then allow higher correlations (Kebbell *et al.*, 1996). Further examination of the data for easy and moderate items suggested that this effect was also due to participants who were *not* confident in their correct answers (because they were guessing correctly), subsequently maintaining this low confidence, but producing less correct responses over time. However, the converse applied for difficult items, where the C-A correlation was initially negative. That is to say, those who were confident in incorrect answers tended to maintain this confidence over time, whilst forgetting their correct answers over time.

Of course, there are numerous other types of interviewing conditions and procedures within the Criminal Justice System that also require investigating over time; i.e. variations of immediate police interview; delayed police interview; examination-in-chief; re-examination; opening and/or closing argument, and so on. The effects of these variations on eyewitness confidence and accuracy were obviously beyond the scope of this thesis. Notwithstanding these limitations, the present results suggest that the intuitive assumption that 'any delay can only be negative' is open to challenge. Instead the effects depend critically on what is being recalled and how it is measured. Both accuracy and confidence do, indeed, decline over time, however, within subjects C-A correlations are relatively unaffected. On the other hand,

between subjects C-A correlations for easy and moderate items tend to improve over time; though those for difficult items are made rather poorer. As the relationship between confidence and accuracy was differentially affected perhaps the approach advisable for future work might be to aim to identify an optimum time delay; that is, the amount of time necessary that might result in an acceptable trade-off between inaccurate responses, and the positive effect of memory consolidation that results in significant between-subjects C-A correlations.

13.3 The effects of cross-examination questioning styles on jurors' assessments of witness confidence and accuracy?

The aim of the previous experiments was to assess objectively the effects of questioning styles and techniques used in the courtroom on witness accuracy and confidence; including the provision of procedural delay. However, as noted in Chapter 11, given that cross-examination is considered to be central to the English Legal System, perhaps more significant in the context of court decision making processes are the effects of questioning style on the ways in which jurors discriminate between those witnesses who are accurate than those who are not. Moreover, given that cross-examination questioning styles do appear to have some effects upon witness performance, the possibility that such styles might influence jurors' perceptions of witnesses' confidence and accuracy seemed the next logical step to take. The aim of Experiment 5 therefore, was to assess jurors' perceptions of eyewitness confidence and accuracy as a function of the type of questioning style used.

One possible hypothesis was that, by suggesting the witness might be wrong, the lawyerese with negative feedback condition might serve to make

observers doubt the accuracy of a given witness' testimony. For instance, Chapter 5 highlighted some research that has shown when mock-jurors hear inconsistent recall testimony they perceive the eyewitness to be less accurate and credible (Berman, Narby & Cutler, 1995; Brewer, Potter, Fisher, Bond & Luszcz, 1999). One way inconsistency could operate to lower perceived credibility might be to lower the perceived confidence in the witness. The latter may be particularly important given that, as noted earlier, if the witness appears to be confident, he or she will be considered more accurate (Brigham & Wolfskeil, 1983; Cutler & Penrod, 1988; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Stuve, 1988; Deffenbacher & Loftus, 1982; Fox & Walters, 1986; Kassin, Rigby and Castillo, 1991; Leippe, Manion, & Romanczyk, 1992; Lindsay, 1994; Lindsay, Wells & O'Connor, 1989; Lindsay, Wells & Rumpel, 1981; Sporer, 1993; Wells, Ferguson & Lindsay, 1981; Wells, Lindsay & Ferguson, 1979; Whitely & Greenberg, 1986).

The results of Experiment 5 indicate that observers seemed able to discriminate to some extent between the accurate and inaccurate witness accuracy in the control and lawyerese conditions but not in the lawyerese with negative feedback condition. Moreover, as they really only had confidence judgments to go on it seems reasonable to assume that the discriminations were based primarily on differences in perceived confidence. The results of Experiment 5 suggest, therefore, that the subtle negative feedback style of questioning destroys the link that observers make between witness confidence and accuracy such that the confident witness is no longer necessarily assumed to be the most accurate.

However, as noted previously, it is somewhat ironic that in Experiment 2, the only condition to show an overall positive significant between subjects correlation between confidence and accuracy (indicating that a confident witness will be a more accurate witness) was the lawyerese with negative feedback condition. In other words, in the one interviewing condition in which accuracy might reasonably have been inferred from confidence, the interviewing style negated such an inference by observers. One implication of this is that, although observers may judge accuracy on the basis of confidence, accuracy judgments are further moderated by style of questioning, such that a confident witness in the negative feedback condition will not necessarily be assumed to be more accurate.

Moreover, Experiment 5 showed that when the poor/less confident witness came first, the perceptions of the witness' confidence and accuracy were inflated. Judgments of witness competence and stress were affected in much the same way; when the poor witness was presented first, perceptions of the poor witness' competence also increased, and perceptions of stress decreased. With regard to perceptions of stress and witness competence, however, there were also trends, significant in the former, for the good witness to be affected in the opposite direction (stress up, competence down) when the poor witness came first. There is no obvious explanation for this, though as noted in Chapter 11, perhaps the failure to establish clear anchors when the poor witness came first might have confused observers somewhat so they might have adopted a 'go for the mean' strategy in these later ratings. One obvious implication of this is that jurors should be instructed to listen to a range of witnesses before making relative judgments as to their accuracy.

No significant effects were found for fairness in Experiment 5. This suggests, perhaps, that lawyerese questioning styles in themselves (i.e. not coupled, for example, with aggressive intonation or nonverbal gestures) may possess sufficient subtlety not to be perceived as intimidatory and therefore unfair. This raises the issue of whether the styles of questioning used here accurately reflected courtroom practice. In the courtroom, during cross-examination, lawyers do frequently ask their questions whilst adopting an aggressive and intimidatory manner; however, the purpose of the present thesis was primarily to investigate and isolate the effects of phrasing; the effects of presentation are an entirely different matter (which is, of course, well worthy of investigation in its own right).

Notwithstanding their limitations, therefore, the present findings suggest that it is no longer sufficient to assume that the styles of examination employed by cross-examiners have no effect on jurors' judgments of witnesses' performance.

13.4 Predicting who makes a good and poor witness under cross-examination

The literature and discussion in Chapter 3 indicated that witnesses are likely to differ in their responses to procedures inherent in cross-examination. In Study 6 questionnaire data collected during the second and fourth studies were analysed in order to examine possible predictors of witness credibility in court.

One measure used was the FNE scale, which focuses primarily on people's concerns with interpersonal evaluation (see Chapter 12 for details). It was predicted that people who score high on the FNE might be prone to become apprehensive in courtroom examination situations and attempt to

avoid social disapproval. Consequently they might be more vulnerable to leading questions (Gudjonsson, 1992), and have detrimental effects on the relationship between confidence and accuracy. Additionally, the Distress subscale of the SAD scale (Watson and Friend, 1969), and a Discomfort scale developed by the present author were administered on the grounds that subjective feelings of distress about such an interview, and the discomfort a witness may feel during this kind of interrogative situation might also influence cognitive performance. For instance, as noted in Chapter 3, 79% of Kassin *et al*'s (1989) experts agreed that; "Very high levels of stress impair the accuracy of eyewitness testimony", and 71% said that the statement was sufficiently reliable enough to offer it in court. In contrast it can be noted that Nolan and Markham (1998) found that highly anxious participants showed greater positive confidence-accuracy correlations; they suggested that this occurred because highly anxious people were more likely to carefully appraise their performance.

Overall, however, the results from Experiment 6 showed no significant relationships for the FNE scale, the Distress subscale of SAD, and Discomfort scale, with witness performance in terms of accuracy and confidence. Clearly with this particular sample any such anxieties were not sufficient to influence these outcomes.

There could be several reasons for this. For instance, it could be argued that the characteristics of the samples used were relatively homogeneous with respect to influence from situational stress. That is to say, in the main, many of the participant witnesses were members of a University; i.e. people who are used to having their knowledge questioned, are reasonably

confident in their abilities and open to criticism, and who have perhaps spent time, through their experiences, working on strategies to deal with stress when performing cognitive tasks. Alternatively, the individual self-rating scales used were not specific to the task in question (i.e. cross-examination), and might therefore have been relatively insensitive. On the other hand, it is possible that, particular concerns about evaluation, distress and discomfort are to be differentiated from general task anxiety (Nolan and Markham, 1998), and are not influential factors in this kind of situation.

According to Gudjonsson's model of interrogative suggestibility, another factor that may influence susceptibility to leading questions is the processing capacity of individuals. So far, however, little research has been conducted on this issue; consequently, Study 6 also examined cognitive performance in the courtroom and neuropsychological correlates of processing capacity.

In Chapter 12 it was emphasised that good confidence and accuracy relationships require careful monitoring and calibration of memory; a job that would seem to imply activation of the frontal lobes (Troyer, 2000; Troyer, Moscovitch & Winocur, 1997; Troyer, Moscovitch, Winocur, Alexander & Stuss, 1998). To assess frontal lobe performance, therefore, measures of phonemic fluency were also administered (see Chapter 12 for full details). If good confidence-accuracy relationships require substantial frontal activation, one might predict that high scorers on the switching component of verbal fluency would show the highest accuracy, and confidence-accuracy correlations.

The analyses in Study 6 indicated that the switching component of verbal fluency, and correspondingly, a capacity for frontal lobe activation, was related to increased positive confidence-accuracy relationships in the short-term, and greater accuracy (in terms of correct answers) in the long-term. In the long-term, however, it seems that, although the accuracy of those high on fluency remains high relative to those low on fluency, the C-A trend is modified such that confidence remains high, regardless of accuracy. As argued in Chapter 12, these findings may simply reflect a statistical artefact because of the large number of analyses involved, however, alternatively they might indicate an interaction between two behavioural factors in those high on verbal fluency; i.e. an ability to remember accurately and calibrate accuracy with confidence, and a general confidence in their own ability. In the short-term, the ability to calibrate accurately might be sufficient to take precedence over the tendency to be overconfident. Over time, however, the overconfidence becomes more prominent.

The reasons for these findings are not obvious. However, one possible explanation is in terms of task difficulty; i.e. the benefits of high frontal function are only shown when the task becomes difficult and executive processing is necessary. In the short term, therefore, high frontal function may have little effect on an easy task, accuracy, but exert an effect on the ability to calibrate accurately, on one more difficult. After a delay, however, accuracy becomes difficult, and high frontal function is an advantage. However, C-A relationships become proportionately more difficult to calibrate accurately, to the extent that even those of high frontal function are not able to do the task very successfully; instead, because perhaps they are used to doing well in

various situations, they become overconfident. It has been noted, for example, that good frontal function may have a number of advantages in everyday behaviour; for example, the ability to plan behaviour, respond to novel circumstances, and inhibit inappropriate responses (Banich, 1997).

This draws attention to the fact that, although much is known about the behaviour of patients with frontal lobe damage who score poorly on verbal fluency (Troyer *et al.*, 1997; 1998), little is known about the personality and social characteristics of those scoring high on verbal fluency within the normal population. For example, in a recent unpublished study by Wagstaff at Liverpool, a significant correlation of .66 (n = 30) was found between FAS switching scores and Synder's Self-Monitoring Scale (1974; 1979) which measures an ability to adapt one's behaviour to different social circumstances so as to present oneself in a favourable light. This might account for the fact that FAS scores were associated with high confidence; perhaps because high FAS scorers were generally more sure of their own abilities.

In general, these results suggest that ordinary and willing witnesses might well differ in their discriminative abilities, and that these abilities might in fact be related to specific cognitive processing capacities that rely heavily on frontal lobe activation. However, in terms of predicting courtroom performance, the present results suggest that possessing high verbal fluency ability may be a mixed blessing. At short intervals it may be useful, aiding confidence-accuracy relationships; after an interval, however, the result may be overconfidence. Nevertheless, after a delay those high on verbal fluency are also the most likely produce correct answers.

In line with previous comments made in Chapter 12, it is important to note that the method might have been more sensitive if the verbal fluency test had been given again prior to the second interview. In retrospect this would probably have been advisable; in fact, surprisingly, no reliability data are available for phonemic fluency; the assumption in clinical work seems to be that that frontal lobe capacity is fairly invariant in the short-term. Reliability data should thus be collected in future work.

13.5 Summary of the main findings

The main findings of the present thesis can, therefore, be summarized as follows.

1. Lawyerese questioning with subtle negative feedback has a significant detrimental effect on witness accuracy to critical items.

2. Standard lawyerese questioning alone (i.e. not including subtle negative feedback) significantly reduces positive within subjects C-A relationships, due primarily to a reduction in confidence in correct answers.

3. For standard lawyerese questioning, a delay of 6 months significantly reduces accuracy for all levels of item difficulty, and confidence for all levels of item difficulty, except for easy items.

4. With regard to point 3, overall, witnesses' own C-A calibrations remain non-significant and unaffected by the delay, while between subjects C-A relationships become significant and positive after a delay, except for difficult items. For difficult items, however, the trend is for non-significant C-A relationships to persist, in fact become slightly worse, over time.

5. Juror/observers tend to link witness confidence with accuracy, however, lawyerese questioning with subtle negative feedback destroys this link.

6. Presenting a poor witness first increases observer's perceptions of witness accuracy, confidence and competence; and decreases perceptions of witness stress.

7. Frontal lobe function is positively related to C-A relationships in the short term, and accuracy after 6 months. After 6 months, however, frontal lobe function is also related to high confidence regardless of accuracy.

13.6 Limitations, future research, and possible recommendations

Several limitations of the research contained within this thesis have been noted which might have implications for future research. It may, therefore, be useful to summarize and re-emphasize them at this point.

One particular limitation of the work with respect to lawyerese questioning styles is the characteristics of the samples used. In real courtroom situations, the lawyers are not simply concerned with getting the maximum information out of willing witnesses, but also 'catching out' those who may deliberately be attempting to hide or distort information. The work presented here with regard to lawyerese questioning does not seem to support the general idea that these kinds of questioning styles might help to maximise information or calibrate witnesses' memories, as has been asserted by the legal profession. In this sense, it is very relevant to practice in the courtroom. However, how effective they are at extracting information from an unwilling and untruthful witness remains to be established.

The fact that all of the participants in the reported research were 'willing' witnesses raises the more general question of ecological validity. For example, in addition to witnesses being 'willing', participants whose mean age was in the mid-twenties, may not necessarily have been representative of witnesses who are called to give evidence in court. It is likely, for example, that witnesses called to give evidence in the courtroom would have a mean age higher than that of those participants used in the studies here.

There are other points worthy of discussion. For example, it could be suggested that experimental models of cross-examination are not likely to create the context in which a real witness gives evidence. There may be problems associated with experimenter bias and demand characteristics such that the environment and set-up might lead to witnesses agreeing with presupposition in a question because they believe that is what the experimenter wants. However, whilst this might be valid, it seems also to be a

strength; that is, this is precisely what examiners wish to achieve when conducting a cross-examination. Indeed, during debrief sessions with individual witnesses, that followed the interview, a number of witnesses commented that they had in fact given evidence in court. These witnesses stated that the questioning in the present studies had recreated this context for them; accordingly, in many cases the witnesses became visibly more tense and nervous as the interview progressed. This suggests that the questioning procedures here might not have been as divorced from real-life contexts as at first on might think.

Further criticism in relation to ecological validity could be levied at the stimulus materials; these were in the form of a videotape recording of a criminal event rather than one which is staged and observed in real-time. However, the event contained within the video stimulus was staged and specifically devised to create an element of realism. This was done by filming the event from the perspective of the witness so as to create, at least, a visual context for the witness whom observed the event.

Most of these major criticisms might have been overcome in an experiment that was proposed as a finale to the work. A trainee barrister was to conduct the questioning scripts in a moot courtroom at the University of Liverpool. A judge and courtroom staff (suitably attired) were to be included in the study specifically to create a more valid courtroom context for the participant witnesses, together with 12 'reasonable' persons (i.e. jurors). The study was set up to run, but during the final planning stages the barrister who had agreed to conduct the study was called to an Inns of Court in London. Arguably, such a study is necessary to move the research context from one

that has some limitations in terms of ecological validity to one that could be considered to be truly applied to the courtroom (as far as possible) whilst still retaining the rigour of experimental design and analysis.

However, there may also be some merit in attempting to differentiate between willing and potentially truthful witnesses and others. For example, it is possible to categorise witnesses into those removed from the defendant who have no apparent motive for maintaining the innocence or guilt of the defendant, and others who may have motives for omitting or distorting information. In the case of the former, unhelpful techniques in eliciting accurate testimony, such as leading, pre-supposition, and feedback strategies could be omitted and replaced with more appropriate techniques.

Witnesses who are called to provide expert opinion on a matter relevant to a given case represent an exemplar of a willing witness. Clearly, the expert is (in principle at least) removed from the defendant and offers the court invaluable opinion for consideration in the context of all other evidence. In terms of the administration of justice, it is difficult to see what is to be gained by a cross-examiner's attempt to confuse and mislead experts into incredulity. Given that the research here indicates that cross-examination strategies can be unhelpful, it seems appropriate for a list of working guidelines for the cross-examination of expert witnesses to be constructed. One possible way of proceeding in this direction might be to incorporate some of lessons learned from the examination of police interviews in Chapter 4. For example, interruption in courtroom examination is often adopted as a deliberate tactic, to prevent witnesses from providing information detrimental to a case; it may, however, be quite obstructive in terms of finding 'the truth'.

It could be suggested, therefore, that, during cross-examination excessive interruption should be avoided. More importantly, however, in principle, cross-examiners should perhaps be discouraged from using leading and pre-suppositional questions, excessive use of question and answer format, and inappropriate sequencing of questions.

Another direction for future research concerns differentiating between the types of negative feedback available for use in interrogative situations; i.e. overt and subtle. Future research should clearly include this differentiation. Moreover, it is also important to assess the effects of overt and subtle feedback upon those witnesses who might be classed as particularly vulnerable to this kind of interviewing (i.e. child; learning disabled; older adults).

The work here did not find any differential effects of a six months delay on within subjects C-A correlations when a formal review of previous judgments was not included. Arguably, following a relatively long delay of 6 months, witnesses would realistically only have memory traces as pointers to where information might be found in response to a specific question, rather than a solid benchmark to which they had previously alluded to. Such an anchor might thus affect any subsequent confidence ratings made. Future work should, therefore, investigate the effects of formal reviews of statements made by witnesses, over varying time delays, to examine effects of these factors upon accuracy and confidence, and C-A relationships.

There are other aspects that pertain to interviewing conditions and procedures within the Criminal Justice System that also require investigating over time; i.e. variations of immediate police interview; delayed police interview; examination-in-chief; cross-examination; re-examination; opening

and closing argument, etc. The effects of these variations on eyewitness confidence and accuracy were obviously beyond the scope of this thesis. The findings, however, relating to between subjects C-A relationships (which increase for easy and moderate items over time) raise the issue of the efficacy of an *overly expedient* system in relation to courtroom appearances for witnesses. Arguably, the Home Office should allow for a full investigation of the impact of delay rather than following the intuitive assumption that 'any delay can only be negative'.

In relation to jurors' perceptions of witnesses, the present research indicates that even although observers may judge accuracy on the basis of confidence, accuracy judgments are further moderated by style of questioning, such that a confident witness in the negative feedback condition will not necessarily be assumed to be more accurate. However, in the courtroom, during cross-examination, lawyers frequently ask their questions whilst adopting an aggressive and intimidatory manner. As the main focus of this thesis was to isolate the effects of phrasing, issues pertaining to the effects of presentation were not addressed. Such issues are certainly worthy of investigation in their own right given that the courts can no longer assume that cross-examination questioning styles have no effect on jurors' judgments of witnesses.

Moreover, order effects were prevalent; such that when the poor/less confident witness came first, the perceptions of the witness' confidence and accuracy were inflated. Again, when the poor witness was presented first; perceptions of witness competence increased and stress decreased. Interestingly, however, there were also trends (significant for stress), for the

good witness to be affected in the opposite direction (stress up, competence down) when the poor witness came first. On the basis of the above order effects, it seems advisable to instruct jurors to listen to a range of witnesses before making relative judgments as to their accuracy.

The possibility that the results from the individual differences data relating to the FAS might indicate the interaction of two behavioural factors in those high on verbal fluency; i.e. an ability to remember accurately and calibrate accuracy with confidence, and a general confidence in their own ability, is appealing, and obviously requires further investigation. Thus, future work might look at the validation of 'switching' as a discriminatory function in witnesses.

To reiterate, in the short-term, the ability to calibrate accurately might be sufficient to take precedence over the tendency to be overconfident; whereas overconfidence might become more prominent later. As noted above, the method would probably have been more sensitive had the verbal fluency test been given again prior to the second interview; however, as verbal fluency has been used mainly in clinical work, no reliability data are available. Such reliability data could be important, and future work might attempt to illuminate the literature in this respect.

13.7 CONCLUSION

With regard to cross-examination, one of the most famous statements made by Wigmore is that:

> "…. it is beyond any doubt the greatest legal engine ever invented for the discovery of the truth …. The fact of this unique and irresistible power remains, and is the reason for our faith in its merits …. cross-examination, not trial by jury, is the great and permanent contribution of the Anglo-American system of law to improved methods of trial procedure" (1974, p.).

If the results of the present thesis have any generalisability, they suggest that such pronouncements about the efficacy of cross-examination procedures cannot be taken on faith, and require vigorous testing.

REFERENCES

Ainsworth, P.B. (1995). Psychology and Policing in a Changing World. Chichester: John Wiley & Sons.

Alden, L. (1989). Short-term structured treatment for avoidant personality disorder. Journal of Consulting and Clinical Psychology, 57, 756-764.

Anderson, J.R., & Bower, G.H. (1973). Human Associative Memory. Washington D.C.: Wigston.

Aschermann, E., Mantwill, M., & Kohnken, G. (1991). An independent replication of the effectiveness of the cognitive interview. Applied Cognitive Psychology, 5, 489-495.

Atkinson, J.M., & Drew, P. (1979). Order in Court. Atlantic Highlands, N.J.: Humanities Press.

Banich, M.T. (1997). Neuropsychology: The neural bases of mental function. New York: Houghton Mifflin.

Baldwin, S. (1987). Cross-Examination: The old rules are still the good ones. TRIAL, 23(7), 76-81.

Bartol, C., & Bartol, A. (1998). History of Forensic Psychology. In A.K. Hess, & I.B. Weiner (Eds.), The Handbook of Forensic Psychology, 2nd Edn. 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. (1994). The fate of errors produced under the Cognitive Interview. Paper presented at the Practical Aspects of Memory Conference, Maryland, USA, July.

Bekerian, D.A., Dennett, J.L., Reeder, C., Sloper, K, Saunders, H., & Evans, L. (1994). The influence of the cognitive interview technique on the attributions of memory. Poster presented at the BPS Annual Conference, Brighton, UK.

Berman, G.L., Narby, D.J., & Cutler, B.L. (1995). Effects of inconsistent statements on mock jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. Law and Human Behavior, 19, 79-88.

Berman, G.L., & Cutler, B.L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. Journal of Applied Psychology, 81(2), 170-177.

Binet, A. (1900). La Suggestibilite. Paris: Doin & Fils.

Binet, A. (1905). La science du temoignage. Annee Psychologique, 11, 128-136.

Bohannon, J.N. (1992). Arousal and memory: Quantity and consistency over the years. In E. Winograd & U. Neisser (Eds.), Affect and accuracy in recall: The problem of "flashbulb" memories, (pp. 65-91). New York: Cambridge University Press.

Boon, J., & Noon, E. (1994). Changing perspectives in cognitive interviewing. Psychology, Crime and Law, 1, 59-69.

Borkovec, T.D., Robinson, E., Pruzinsky, T., & DePree, J.A. (1983). Preliminary exploration of worry: Some characteristics and processes. Behaviour Research and Therapy, 21, 9-16.

Bothwell, R.K., Deffenbacher, K.A. & Brigham, J.C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. Journal of Applied Psychology, 72, 691-695.

Bourne, L.E., Dominowski, R.L., & Loftus, E.F. (1979). Cognitive processes. Englewood Cliffs, New Jersey: Prentice-Hall.

Bower, G. H. (1967). A multi-component theory of memory trace. In K.W. Spence & J.T. Spence (Eds.) The Psychology of Learning and Motivation, Vol 1, (pp. 229-325). New York: Academic Press.

Brennan, M. (1995). The discourse of denial: Cross-examining child victim witnesses. Special issue: Laying down the law: Discourse analysis of legal institutions. Journal of Pragmatics, 23, 71-91.

Brewer, N., Potter, R., Fisher, R.P., Bond, N., & Luszcz, M.A. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. Applied Cognitive Psychology, 13(4), 297-313.

Brigham, J.C., & Wolfskeil, M.P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. Law and Human Behaviour, 7, 337-349.

Britt, T.W. (1992). The self-consciousness scale: On the stability of the three-factor structure. Personality and Social Psychology Bulletin, 18, 748-755.

Brown, R.A. (1973). A first language: the early stages. Cambridge, Mass.: Harvard University Press.

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. Memory and Cognition, 20, 277-290.

Cady, H.M. (1924). On the psychology of testimony. American Journal of Psychology, 35, 110-112.

Carr, T.H., Deffenbacher, K.A., & Leu, J.R. (1979). Is there less interference in memory for faces? Paper presented at the meeting of the Psychonomic Society, Phoenix, Arizona.

Christianson, S-A. (1984).The relationship between induced emotional arousal and amnesia. Scandianavian Journal of Psychology, 25, 147-160.

Christianson, S.-A. (1992). Emotional Stress and Eyewitness Memory: A Critical Review. Psychological Bulletin, 112(2), 284-309.

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.

Christianson, S.-A., & Loftus, E.F. (1987). Memory for traumatic events. Applied Cognitive Psychology, 1, 225-239.

Christianson, S.-A., & Nilsson, L.-G. (1984). Functional amnesia as induced by a psychological trauma. Memory and Cognition, 12, 142-155.

Clifford, B.R., & George, R. (1996). A Field Evaluation of Training in Three Methods of Witness/Victim Investigative Interviewing. Psychology, Crime & Law, 2, 231-248.

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(3), 364-370.

Clifford, B.R., & Scott, J. (1978). Individual and Situational Factors in Eyewitness Testimony. Journal of Applied Psychology, 63(3), 352-359.

Coulthard, M. (1981). Developing the description. In M. Coulthard & M. Montgomery (Eds.), Studies in Discourse Analysis, (pp. 13-31). London: Routledge & Kegan Paul.

Croft, S. (1995). Helping victims to remember. Police, 28,13-14.

Cross, Sir, R., Wilkins, N., & Tapper, C. (1986). An outline of the law of evidence. London: Butterworth & Co.

Cutler, B.L., & Penrod, S.D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. Journal of applied Psychology, 73, 281-290.

Cutler, B.L., & Penrod, S.D. (1995). Mistaken Identification: The Eyewitness, Psychology and the Law. New York: Cambridge University Press.

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., & Stuve, T.E. (1988). Juror decision making in eyewitness identification cases. Law and Human Behaviour, 12, 41-55.

Danet, B. (1980). Language in the legal process. Law & Society Review, 14, 445-564.

Danet, B., & Bogoch, B. (1980). Fixed fight or free-for all? An empirical study of combativeness in the adversary system of justice. British Journal of Law and Society, 7, 36-60.

Davies, G.M., Westcott, H.L., & Horan, H. (2000). The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. Psychology, Crime & Law, 6(2), 81-97.

Davis, P.J. (1990). Repression and the inaccessibility of emotional memories. In J.L. Singer (Ed.), Repression and dissociation: Implications for personality theory, psychopathology, and health, (pp.387-403). Chicago: University of Chicago Press.

Deffenbacher, K.A. (1980). Eyewitness accuracy and confidence. Law and Human Behaviour, 4, 243-260.

Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behaviour? Law and Human Behavior, 6, 15-30.

Deffenbacher, K.A., Carr, T.H., & Leu, J.R. (1981). Memory for words, pictures and faces: Retroactive interference, forgetting and reminiscence. Journal of Experimental Psychology: Human Learning and Memory, 7, 299-305.

Dent, H.R. (1978). Interviewing child witnesses. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.). Practical Aspects of Memory: Current Research and Issues. New York: John Wiley & Sons.

Dent, H., & Stephenson, G. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. British Journal of Social and Clinical Psychology, 18, 41-51.

Dodd, D.H., & Bradshaw, J.M. (1980). Leading questions and memory: Pragmatic constraints. Journal of Learning and Verbal Behavior, 19, 695-704.

Drew, P. (1990). Strategies in the contest between lawyer and witness in cross-examination. In J.N. Levi & A. Graffam Walker (Eds.). Language in the Judicial Process, (39-64). New York: Plenum Press.

Du Cann, R. (1964). The Art of the Advocate. Harmondsworth: Penguin.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. New York: Columbia Teacher's College, (Original work published 1885).

Eggleston, Sir, R. (1978). Evidence, Proof and Probability. London: Weidenfield & Nicolson.

Ellison, L. (1998). Cross-Examination in Rape Trials. The Criminal Law Review, 605-615.

Erickson, B.E., Lind, A., Johnson, B.C., & O'Barr, W.M. (1987). Speech style and impression formation in a court setting: the effects of 'powerful' and 'powerless' speech. Journal of Experimental Social Psychology, 14, 266-279.

Evans, F.J. (1967). Suggestibility in the normal waking state. Psychological Bulletin, 67, 114-129.

Evans, F.J. (1989). The independence of suggestibility, placebo response, and hypnotizability. In V.A. Gheorghiu., P. Netter., H.J. Eysenck., & R. Rosenthal (Eds.). Suggestion and Suggestibility: Theory and Research, (pp.145-154). London: Springer-Verlag.

Evans, K. (1995). Advocacy in Court: A Beginner's Guide, 2nd Edn. London: Blackstone.

Eysenck, M.W. (1947). Dimensions of Personality. London: Routledge & Kegan Paul.

Fillmore, C.J. (1969). Types of lexical information. In F. Fiefer (Ed.), Studies in Syntax and Semantics, (pp. 120-121). Dordrecht: D. Reidel.

Fisher, R.P. (1995). Interviewing victims and witnesses of crime. Psychology, Public Policy, and Law, 1(4), 732-764.

Fisher, R.P., & Cutler, B.L. (1996). The relation between consistency and accuracy of eyewitness testimony. In G. Davies, S. Lloyd-Bostock, M. McMurran & C. Wilson (Eds.). Psychology, Law, and Criminal Justice, (pp. 21-28). Berlin, Germany: Walter de Gruyter.

Fisher, R.P., & Geiselman, R.E. (1992). Memory-Enhancing Techniques for Investigative Interviewing: The Cognitive Interview. Illinois: Charles C. Thomas.

Fisher, R.P., & Price-Roush, J. (1986). Question order and eyewitness memory. Unpublished manuscript, Department of Psychology, Florida International University, USA.

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(5), 722-727.

Fisher, R.P., Geiselman, E., & Raymond, D.S. (1987). Critical Analysis of Police Interview Techniques. Journal of Police Science and Administration, 15(3), 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 and Administration, 15, 291-297.

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.

Frege, G. (1975). Uber Sinn und Bedeutung. In Funktion, Begriff, Bedeutung: Funf logische Studien, pp. 40-65, first published in 1892. Translated in P. Geach & M. Black (1980), Translations from the Philosophical Writings of Gottlob Frege, (3rd Edn., pp. 56-78). Oxford: Blackwell.

Fruzzetti, A.E., Toland, K., Teller, S.A., & Loftus, E.F. (1992). Memory and eyewitness testimony. In M.M. Gruneberg & P.E. Morris (Eds.). Practical Aspects of Memory: Current Research and Issues, Vol. 1, (pp. 18-50). London: Routledge.

Geiselman, R.E. (1996). On the use and efficacy of the cognitive interview. Psycoloquy, 7(6),witness-memory.2.geiselman.
Located at www.ai.univie.at/archives/Psycoloquy/1996.V7/0016.html

Geiselman, R.E., Fisher, R.P., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness response to leading and misleading questions under the cognitive interview. Journal of Police Science and Administration, 14, 31-39.

Geiselman, R.E., Fisher, R.P., Firstenberg, I., Hutton, L.A., Sullivan, S., Avetissian, I., & Prosk, A. (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 mneumonics versus hypnosis. Journal of Applied Psychology, 70, 401-412.

Geiselman, R.E., Fisher, R.P., MacKinnon, D.P., & Holland, H.L. (1986). Enhancement of eyewitness memory with the cognitive interview. American Journal of Psychology, 99, 385-401.

George, R. (1991). A field and experimental evaluation of three methods of interviewing eyewitnesses and victims of crime. Unpublished Masters Thesis, Polytechnic of East London, London, UK.

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

Glissan, J.L. (1991). Cross-examination Practice and Procedure: An Australian Perspective. Sydney: Butterworths.

Goldstein, A.F., Chance, J.E., & Schneller, G.R. (1989). Frequency of eyewitness identification in criminal cases. Bulletin of the Psychonomic Society, 27, 71-74.

Goodman, G.S., Hepps, D.G., & Reed, R.S. (1986). The child victim's testimony. In A. Haralamic (Ed.), New issues for child advocates, (pp. 167-176). Phoenix, AZ: Arizona Council of Attorneys for Children.

Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. Journal of Applied Psychology, 82(4), 599-613.

Greene, E., Flynn, M.S., & Loftus, E.F. (1982). Inducing resistance to misleading information. Journal of Verbal Learning and Verbal Behaviour, 21, 207-219.

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

Grigg-Spall, I., & Ireland, P. (1992). The Critical Lawyers' Handbook. London: Pluto Press.

Gruneberg M.M., & Sykes, R.N. (1993). The generalisability of confidence-accuracy studies in eyewitnessing. Memory, 1, 185-189.

Gudjonsson, G.H. (1987). Historical background to suggestibility: how interrogative suggestibility differs from other types of suggestibility. Personality and Individual Differences, 8, 347-355.

Gudjonsson, G.H. (1988). Interrogative suggestibility: Its relationship with assertiveness, social-evaluative anxiety, state anxiety and method of coping. British Journal of Clinical Psychology, 27, 159-166.

Gudjonsson, G.H. (1989a). Theoretical and empirical aspects of suggestibility. In V.A. Gheorghiu., P, Netter., H.J. Eysenck., & R. Rosenthal (Eds.). Suggestion and Suggestibility: Theory and Research, pp. 135-143. London: Springer-Verlag.

Gudjonsson, G.H. (1989b). The effects of suspiciousness and anger on suggestibility. Medicine, Science and the Law, 29(3), 229-232.

Gudjonsson, G.H. (1992). The Psychology of Interrogations, Confessions and Testimony. Chichester: John Wiley & Sons.

Gudjonsson, G.H., & Clark, N.K. (1986). Suggestibility in police interrogation: a social psychological model. Social Behaviour, 1, 83-104.

Gwyer, P., & Clifford, B.R. (1997). The effects of the cognitive interview on recall, identification, confidence, and the confidence-accuracy relationship. Applied Cognitive Psychology, 11, 121-145.

Hall, D.F. (1980, May). Memory for faces and words: Effects of stimulus presentation interval and depth of processing. Paper presented at the meeting of the Midwestern Psychological Association, St. Louis.

Harris, S. (1984). Questions as a mode of control in magistrates' courts. International Journal of Society and Language, 49, 5-27.

Hastie, R., Landsman, R., & Loftus, E.F. (1978). Eyewitness testimony: The dangers of guessing. Jurimetrics Journal, 19, 1-8.

Havatny, N. & Strack, F. (1980). The impact of a discredited key witness. Journal of Applied Social Psychology, 10, 490-509.

Heath, W.P., & Erickson, J.R. (1998). Memory for central and peripheral actions and props after varied post-event presentation. Legal and Criminological Psychology, 3, 321-346.

Hertel, P.T., Cosden, M., & Johnson, P.J. (1980). Passage recall: schema change and cognitive flexibility. Journal of Educational Psychology, 72, 133-140.

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. Memory and Cognition, 18, 496-506.

Hickey, L. (1993). Presupposition Under Cross-Examination. International Journal for the Semiotics of Law, 1(16), 89-109.

Home Office (1999). Listing arrangements for criminal cases in the Magistrates' Courts: A draft protocol. Located at www.cjsonline.org/legal/trials_issue_group.htm

Hull, C. (1933). Hypnosis and Suggestibility. New York: Appleton-Century-Crofts.

Irving, B., & Hilgendorf, L. (1980). Police Interrogation: The Psychological Approach. Research Studies No. 1. London: HMSO.

Jackson, B.S. (1988). Law, fact, and narrative coherence. Merseyside, UK: Deborah Charles Publications.

Jones, W.H., Cheek, J.M., & Briggs, S.R. (Eds.) (1986). Shyness: Perspectives on research and treatment. New York: Plenum.

Kassin, S.M., Ellsworth, P.C., & Smith, V.L. (1989). The 'general acceptance' of psychological research on eyewitness testimony. American Psychologist, 44, 1089-1098.

Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy-confidence correlation in eyewitness testimony: limits and extension of the retrospective self-awareness effect. Journal of Personality and Social Psychology, 61, 698-707.

Kassin, S.M., Tubb, V.A., Hosch, H.M., & Memon, A. (2001). On the 'general acceptance' of eyewitness testimony research: A new survey of the experts. American Psychologist, 56(5), 405-416.

Kebbell, M.R., & Giles, D.C. (2000). Lawyers' questions and witness confidence: Some experimental influences of complicated lawyers' questions on witness confidence and accuracy. The Journal of Psychology, 134(2), 129-139.

Kebbell, M.R., & Deprez, S. (1998). Lawyers' questions and witnesses' answers in court: An evaluation. Paper presentation to the International Congress of Applied Psychology, San Francisco, USA.

Kebbell, M.R., & Johnson, S.D. (2000). Lawyers' questioning: The effect of confusing questions on witness confidence and accuracy. Law and Human Behavior, 24(6), 629-641.

Kebbell, M.R., & Milne, R. (1998). Police officers' perception of eyewitness factors in forensic investigations: A survey. The Journal of Social Psychology, 138, 323-330.

Kebbell, M.R., & Wagstaff. G.F. (1996). Enhancing the practicality of the cognitive interview in forensic situations. Psycoloquy, 7(16), witness-memory.3.kebbell. Located at www.ai.univie.ac.at/archives/Psycoloquy/1996.V7/0022.html

Kebbell, M.R., Wagstaff, G.F., & Covey, A.C. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. British Journal of Psychology, 87, 653-662.

Kennedy, T.D., & Haygood, R.C. (1992). The discrediting effect in eyewitness testimony. Journal of Applied Social Psychology, 22(1), 70-82.

Kohnken, G., Thurer, C. & Zoberbier, D. (1994). The cognitive interview: Are the interviewer's memories enhanced, too? Applied Cognitive Psychology, 8, 13-24.

Kohnken, G., Milne, R., Memon, A., & Bull, R. (1994). A meta-analysis of the effects of the cognitive interview. Paper presented at the Biennial Conference of the American Psychology Law Society, Santa Fe, New Mexico, USA.

Lavrakas, P.J., & Bickman, L. (1975). What makes a good witness? Paper presented at the meeting of the American Psychological Assoication, Chicago, (August-September).

Leary, M.R. (1983a). A brief version of the Fear of Negative Evaluation Scale. Personality and Social Psychology Bulletin, 9, 371-376.

Leary, M.R. (1983b). Understanding social anxiety: Social, personality, and clinical perspectives. Beverly Hills, CA: Sage.

Leary, M.R., Knight,, P.D., & Johnson, K.A. (1987). Social anxiety and dyadic conversation: A verbal response analysis. Journal of Social and Clinical Psychology, 5, 34-50.

Leippe, M.R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eywitness accuracy and confidence. Law and Human Behaviour, 4, 261-274.

Leippe, M.R., Manion, 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.

Leung, K., & Lind, E.A. (1986). Procedural justice and culture: effects of culture, gender and investigator status on procedural preferences. Journal of Personality and Social Psychology, 50, 1134-1140.

Lind, E.A., & Tyler, T.R. (1988). The social psychology of procedural justice. New York: Plenum.

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., Glenn, J., Nosworthy, R.M., & Martynuck, C. (1994). Using mug shots to find suspects. Journal of Applied Psychology, 79(1), 121-130.

Lindsay, R.C.L., Wells, G.L., & O'Connor, 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 across situations? Journal of Applied Psychology, 66, 79-89.

Lipton, J.P. (1977). On the psychology of eyewitness testimony. Journal of Applied Psychology, 62, 90-95.

Lloyd-Bostock, S.M.A., & Clifford, B.R. (1983). Evaluating Witness Evidence: Recent Psychological Research and New Perspectives. Chichester: John Wiley & Sons.

Loftus, E.F. (1974). Reconstructing Memory: Incredible Eyewitness. Psychology Today, 8, 116-119.

Loftus, E.F. (1975). Leading Questions and the Eyewitness Report. Cognitive Psychology, 7, 560-572.

Loftus, E.F. (1979). Eyewitness Testimony. Cambridge, Mass.: Harvard University Press.

Loftus, E.F. (1981). Metamorphosis: alterations in memory produced by mental bonding of new information to old. In J. Long and A. Baddeley (Eds.), Attention and Performance, Vol. IX, (pp. 417-434). Hillsdale, NJ.: Erlbaum.

Loftus, E.F., & Burns, T.E. (1982). Mental shock can produce retrograde amnesia. Memory and Cognition, 10, 318-232.

Loftus, E.F., & Kallman, H.J. (1979). Encoding and use of detail information in picture recognition. Journal of Experimental Psychology: Human Learning and Memory, 5, 197-211.

Loftus, E.F., & Ketcham, K. (1991). Witness for the Defense: The accused, the eyewitness, and the expert who puts memory on trial. St. Martin's Press: New York.

Loftus, E.F., & Palmer, J.C. (1974). Reconstruction of automobile destruction: An example of the 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(1), 86-88.

Loftus, E.F., Altman, D., & Geballe, R. (1975). Effects of questioning upon a witness' later recollections. Journal of Police Science and Administration, 3, 162-165.

Loftus, E.F., Loftus, G.R., & Messo, J. (1987). Some facts about "weapon focus". Law and Human Behaviour, 11, 55-62.

Luus, C.A.E., & Wells, G.L. (1994). Eyewitness identification performance. In D.F. Ross, J.D. Read & M.P. Toglia (Eds.), Adult Eyewitness Testimony: Current trends and developments, (pp. 348-361). New York: Cambridge University Press.

McCloskey, M., Egeth, H., Webb, E., Washburn, A., & McKenna, J. (1981). Eyewitnesses, jurors and the issue of overbelief. Unpublished manuscript, John Hopkins University.

McConville, M., Hodgson, J., Bridges, L., & Pavlonic, A. (1994). Standing Accused: The organisation and practices of criminal lawyers in Britain. Oxford: Clarendon Press.

McDougall, W. (1908). An Introduction to Social Psychology. London: Methuen.

Maass, A., & Kohnken, G. (1989). Eyewitness identification: Simulating the "weapon effect". Law and Human Behaviour, 13, 397-408.

Mauet, T.A., & McCrimmon, L.A. (1993). Fundamentals of Trial Techniques. Australian Edition. Melbourne: Longman.

Mandler, G. (1980). Recognizing: The judgment of a previous occurrence. Psychological Review, 87, 252-271.

Manktelow, K.I., & Over, D.E. (1990a). Inference and understanding: a philosophical and psychological perspective. London: Routledge.

Manktelow, K.I., & Over, D.E. (1990b). Deontic thought and the selection task. In K. Gilhooly, M. Keane, R. Logie, & G. Erdos (Eds.), Lines of thinking: Reflections on the psychology of thought, (Vol 1). Chichester: John Wiley & Sons.

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

Marcuse, F.L. (1976). Hypnosis: Fact and fiction. Harmondsworth: Penguin Books.

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.

Marshall, J., Marquis, K.H., & Oskamp, S. (1971). Effects of kind of question and atmosphere of interrogation on accuracy and completeness of testimony. Harvard Law Review, 84, 1620-1643.

May, R. (1990). Criminal Evidence, 2nd Edn. London: Sweet & Maxwell.

Memon, A., & Bull, R. (1991). The cognitive interview: Its origins, empirical support, evaluation, and practical implications. Journal of Community and Applied Social Psychology, 1, 291-307.

Memon, A., & Kohnken, G. (1992). Helping witnesses to remember more: the cognitive interview. Expert Evidence, 2, 39-48.

Memon, A., & Stevenage, S. (1996). Interviewing eyewitnesses: What works and what doesn't? Psycoloquy. 7(6), witness-memory.1.
Located at www.ai.univie.ac.at/archives/Psycoloquy/1996.V7/0032.html

Memon, A., Bull, R., & Smith, M. (1995). Improving the quality of the police interview: Can training in the use of cognitive techniques help? Policing and Society, 5, 53-68.

Milberg, S., & Clark, M.S. (1988). Moods and compliance. British Journal of Social Psychology, 27, 79-90.

Milgram, S. (1963). Behavioural study of obedience. Journal of Abnormal and Social Psychology, 67, 371-378.

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. Human Relations, 18, 57-76.

Milgram, S (1974). Obedience to Authority. London: Tavistock.

Millar, G.R., & Burgoon, J.K. (1982). Factors affecting assessments of witness credibility. In L.N. Kerr and R.M. Bray (Eds.), The Psychology of the Courtroom, (169-199). San Diego: Academic Press.

Mitchell, K.J., Livosky, M., & Mather, M. (1998). The weapon focus effect revisited: The role of novelty. Legal and Criminological Psychology, 3, 287-303.

Murphy, P. (1994). Evidence & Advocacy, 4th Edn. Blackstone Press Ltd: London.

Muscio, B. (1915). The influence of the form of a question. British Journal of Psychology, 8, 351-386.

Myers, D.G. (1996). Social Psychology, (5th Edn.). New York: McGraw-Hill.

Neisser, U, & Harsch, N. (1990, February). Phantom flashbulbs: False recollections of hearing the news about Challenger. Paper presented at the Emory Cognition Project Conference on Affect and Flashbulb Memories, Atlanta,GA.

Nelson, T.O. (1988). Predictive accuracy of feeling of knowing across different criterion tasks and across different subject populations and individuals. In. M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.). Practical aspects of memory: Current research and issues, Vol. 1, (pp. 190-196). Chichester, England: John Wiley & Sons.

Nisbett, R.E., & Ross, L. (1980). Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs, NJ.: Prentice-Hall.

Nolan, J., & Markham, R. (1998). The accuracy-confidence relationship in an eyewitness task: Anxiety as a modifier. Applied Cognitive Psychology, 12, 43-54.

O' Barr, W.M. (1982). Language, power and strategy in the courtroom. New York: Academic Press.

Oppenheim, A.N. (1966). Questionnaire design and attitude measurement. London: Heinemann.

Orne, M.T., Soskes, D.A., Dinges, D.F., & Orne, E.C. (1984). Hypnotically induced testimony. In G. Wells & E. Loftus (Eds.), Eyewitness testimony: Psychological perspectives, (pp. 171-213). New York: Cambridge University Press.

Oxford Dictionary of Current English (1967). Compiled by F.G. Fowler and H.W. Fowler Fourth Edition, Revised by H.G. Le Mesurier and E. McIntosh, Oxford: Oxford University Press.

Patterson, M.L., & Strauss, M.E. (1972). An examination of the discriminant validity of the social avoidance and distress scale. Journal of Consulting and Clinical Psychology, 39, 1969.

Penrod, S., Loftus, E., & Winkler, J. (1982). The Reliability of Eyewitness Testimony: A Psychological Perspective. In N.L. Kerr (Ed.), The Psychology of the Courtroom, (pp. 119-168). New York: Academic 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(1), 144-147.

Perry, N.W., McAuliff, B.D., Tam, P., Claycomb, L., Dostal, C., & Flanagan, C. (1995). When lawyers question children: Is justice served? Law and Human Behaviour, 19(6), 609-629

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 Researchand Issues, (pp. 89-94). New York: John Wiley & Sons.

Powers, P.A., Andriks, J.L., & Loftus, E.F. (1979). Eyewitness accounts of females and males. Journal of Applied Psychiatry, 64, 339-347.

Pruzinsky, T., & Borkovec, T.D. (1990). Cognitive and personality characteristics of worriers. Behaviour Research and Therapy, 28, 507-512.

R v. Thomson [1912] 3 KB 19.

Read, J.D. (2001). Introduction: Trauma, stress, and autobiographical memory. Special Issue of Applied Cognitive Psychology, 15, 1-5.

Rand Corporation (1975). The criminal investigation process, (Vols 1-3). Rand Corporation Technical Report R-1777. Santa Monica, California: DOJ.

Robinson, M.D., & Johnson, J.T. (1996). Recall memory, recognition memory, and the eyewitness confidence-accuracy correlation. Journal of Applied Psychology, 81, 587-594.

Russell, B. (1905). On denoting. Mind, XIV, 479-493.

Ryan, R.H., & Geiselman, R.E. (1991). Effects of biased information on the relationship between eyewitness confidence and accuracy. Bulletin of the Psychonomic Society, 29, 7-9.

Sanders, G.S. (1986). The usefulness of eyewitness research from the perspective of police investigators. Unpublished manuscript. State university of New York at Albany.

Saunders, D.M., Vidmar, N., & Hewitt, E.C. (1983). Eyewitness testimony and the discrediting effect. In S.M.A. Lloyd-Bostock & B.R. Clifford (Eds.), Evaluating witness evidence, (pp. 57-78). New York: John Wiley & Sons.

Schlenker, B.R., & Leary, M.R. (1982). Social anxiety and self-presentation: A conceptualisation and model. Psychological Bulletin, 92, 641-669.

Schooler, J.W., & Loftus, E.F. (1986). Individual differences and experimentation: Complementary approaches to interrogative suggestibility. Social Behaviour, 1, 105-112.

Schooler, J.W., & Loftus, E.F. (1993). Multiple mechanisms mediate individual differences in eywitness accuracy and suggestibility. In J.M. Pucket & H.W. Reese (Eds.). Mechanisms of everyday cognition, (pp. 177-203). Hillsdale, NJ: Erlbaum.

Sellars, W. (1954). Presupposing. The Philosophical Review, LXIII, 197-215.

Shepherd, E. (1988). Developing interview skills. In P. Southgate (Ed.), New Directions in Police Training. London: HMSO.

Shepherd, E. (1993). Aspects of Police Interviewing. Issues in Criminological and Legal Psychology, 18, 3-4.

Smith, V.L., Kassin, S., & Ellsworth, P.E. (1989). Eyewitness accuracy and confidence: Within-versus between-subjects correlations. Journal of Applied Psychology, 74, 356-359.

Snyder, M. (1974). Self-monitoring of expressive behaviour. Journal of Personality and Social Psychology, 30, 526-537.

Snyder, M. (1979). Self-monitoring processes. In L. Berkovitz (Ed.), Advances in experimental social psychology, Vol. 12, (pp. 85-128). New York: Academic Press.

Sporer, S.L. (1992). Post-dicting eyewitness accuracy: Confidence, decision 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., Penrod, S.D., Read, D., & Culter, B.L. (1995). Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. Psychological Bulletin, 118, 315-327.

Stern, W. (1910). Abstracts of lectures on the psychology of testimony and on the study of individuality. American Journal of Psychology, 21, 273-282.

Stern, W. (1938). General Psychology: From the Personalistic Standpoint. New York: Macmillan.

Stern, W. (1939). The psychology of testimony. Journal of Abnormal and Social Psychology, 34, 3-20.

Stone, M. (1988). Cross-examination in Criminal Trials. London: Butterworths.

Stricker, L.J., Messick, S., & Jackson, D.N. (1967). Suspicion of deception: Implications for conformity research. Journal of Personality and Social Psychology, 5(4), 379-389.

Tooley, V., Brigham, J., Maass, A., & Bothwell, R. (1987). Facial recognition: weapon effect and attentional focus. Journal of Applied Social Psychology, 17, 845-859.

Tousignant, J.P., Hall, D., & Loftus, E.F. (1986). Discrepancy detection and vulnerability to misleading postevent information. Memory and Cognition, 14, 329-338.

Troyer, A.K. (2000). Normative data for clustering and switching on verbal fluency tasks. Journal of Clinical and Experimental Neuropsychology, 22(3), 370-378.

Troyer, A.K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. Neuropsychology, 11(1), 138-146.

Troyer, A.K., Moscovitch, M., Winocur, G., Alexander, M.P., & Stuss, D. (1998). Clustering and switching on verbal fluency: the effects of focal frontal- and temporal-lobe lesions. Neuropsychologia, 36(6), 499-504.

Tulving, E. (1974). Cue-dependent forgetting. American Scientist, 62, 74-82.

Turtle, J.W., & Wells, G.L. (1988). Children versus adults as eyewitnesses: Whose testimony holds up under cross-examination? In M.M. Gruneberg, P.E. Morris, & R.N. Sykes, (Eds.), Practical Aspects of Memory: Current Research and Issues, (pp. 27-33). New York: John Wiley & Sons.

Turtle, J.W., & Yuille, J.C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hyperamnesia. Journal of Applied Psychology, 79, 260-271.

Wagenaar, W.A. (1986). My memory: A study of autobiographical memory over six years. Cognitive Psychology, 18, 225-252.

Wagstaff, G.F. (1982). Helping a Witness Remember: A Project in Forensic Psychology. Police Research Bulletin, 38, 56-58.

Wagstaff, G.F. (1991). Suggestibility: A Social Psychological Approach. In J.F. Schumaker (Ed.), Human Suggestibility: Advances in theory, research and application, (pp. 132-145). London: Routledge.

Wagstaff, G.F. (1999). Hypnotically induced testimony. In A. Heaton-Armstrong, E. Shepard, & D Wolchover (Eds.). Analysing Witness Testimony: A guide for legal practitioners and other professionals, (pp. 162-177). London: Blackstone Press.

Wagstaff, G.F., & Kelhar, S. (1993). On the roles of control and outcomes in procedural justice. Psychological Reports, 73, 121-122.

Walton, D. (1989). Informal Logic: A Handbook for Critical Argumentation. Cambridge: Cambridge University Press.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. Journal of Consulting and Clinical Psychology, 33, 448-457.

Weinberg, H.I., & Baron, R.S. (1982). The discredible eyewitness. Personality and Social Psychology Bulletin, 8, 60-67.

Weingartner, H., & Parker, E.S. (1984). Memory consolidation: Psychobiology of cognition.

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. (1993). What do we know about eyewitness identification? American Psychologist, 48(5), 553-571.

Wells, G.L. & Murray, M. (1984). Eyewitness confidence. In G.L. Wells & E.F. Loftus (Eds.), Eyewitness Testimony: Psychological Perspectives, (pp. 155-170). New York: Cambridge University Press.

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., Lindsay, R.C.L. & Ferguson, T.J. (1979). Accuracy, confidence and juror perceptions in eyewitness testimony. Journal of Applied Psychology, 64, 440-448.

Westcott, H.L. (1995). Children's experiences of being examined and cross-examined: The opportunity to be heard? Expert Evidence, 4(1), 13-19.

Whipple, G.M. (1909). The observer as a reporter: A survey of the psychology of testimony. Psychological Bulletin, 6, 153-170.

Whipple, G.M. (1910). Recent literature on the psychology of testimony. Psychological Bulletin, 7, 365-368.

Whipple, G.M. (1917). Psychology of testimony. Psychological Bulletin, 14, 234-236.

Whitely, B.E., & Greenberg, M.S. (1986). The role of eyewitness confidence in juror perceptions of credibility. Journal of Applied Social Psychology, 16, 387-409.

Wigmore, J.H. (1974). Evidence. Boston: Little Brown.

Wright, D.B., & Stroud, J.N. (1998). Memory quality and misinformation for peripheral and central objects. Legal and Criminological Psychology, 3, 273-286.

Yarmey, A.D., & Jones, H.P.T. (1983). Is the psychology of eyewitness identification a matter of common sense? In S.M.A. Lloyd-Bostock and B.R. Clifford (Eds.), Evaluating Witness Evidence, (pp.13-40). Chichester: John Wiley & Sons.

Yerkes, P.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit formation. Journal of Comp. Neurology and Psychology, 18, 459-482.

Yuille, J.C., & Cutshall, J.L. (1986). A case study of eyewitness memory of a crime. Journal of Applied Psychology, 71(2), 291-301.

Yuille, J.C., & Cutshall, J.L. (1989). Analysis of the statements of victims, witnesses and suspects. In J.C. Yuille (Ed.), Credibility assessment. Norwell, MA: Kluwer Academic.

Zanni, G.R., & Offermann, J.T. (1978). Eyewitness testimony: An exploration of question wording upon recall as a function of neuroticism. Perceptual and Motor Skills, 46, 163-166.

APPENDICES

APPENDIX 1: Control Interview 1

Control	Resp (Y/N)	Conf
1. Were three red cars parked opposite the bus stop?	**Y** N	
2. Did the younger man have gloves on?	Y **N**	
3. Would you say the older man was shaven?	**Y** N	
4. Would you say the driver wore glasses?	**Y** N	
5. Did the older man have a watch on his wrist?	**Y** N	
6. Did the men say anything to each other during the attack?	Y **N**	
7. Did the car driver leave the engine running during the attack?	**Y** N	
8. Did the attacker's car turn out of the junction almost opposite?	Y **N**	
9. Would you say that the getaway car was a saloon?	**Y** N	
10. Did the victim get pushed into the back of the car?	**Y** N	
11. Would you say that two women looked on at the attack?	**Y** N	
12. Did the younger man put a cloth over the victim's mouth?	**Y** N	
13. When the attacker's car left, did a red car follow?	Y **N**	
14. Would you say the younger man was tall?	**Y** N	
15. And would you say he was clean shaved?	**Y** N	
16. Did a bus arrive up the road shortly before the attack?	**Y** N	
17. Did the victim shout out?	**Y** N	
18. Did the older man have a ring on his finger?	**Y** N	
19. Did the attacker's car have four doors?	Y **N**	
20. Would you say the younger man's hair was receding?	Y **N**	
21. Did the victim's bag fall to the ground?	**Y** N	
22. Did a man with a buggy stand and watch the attack from across the road?	Y **N**	
23. Would you say that the victim struggled?	**Y** N	
24. Did another woman post a letter into the yellow box?	Y **N**	
25. Did the woman who walked past the bus stop have two bags?	Y **N**	
26. Would you say both men were wearing dark trousers?	Y **N**	
27. Did the attackers arrive in a silver car?	**Y** N	
28. Did the victim fall to the ground?	Y **N**	
29. Would you say the other woman who checked the timetable had short hair?	Y **N**	
30. Did the younger man have a moustache?	Y **N**	
31. Would you say both men were wearing jeans?	Y **N**	
32. Did the gun have a long barrel?	Y **N**	
33. Did the younger man have glasses on?	Y **N**	
34. Would you say the younger man's shirt was light green?	Y **N**	
35. And would you say the older man's coat was black?	**Y** N	
36. Did the other woman say the time was a quarter-to one?	**Y** N	
37. Did the victim walk toward the car when it pulled up?	Y **N**	
38. Did a mirror fall to the ground during the attack?	**Y** N	
39. Did the gang consist of two men?	Y **N**	
40. Did the older man have baseball shoes on?	Y **N**	
41. Did the older man say anything to the victim?	**Y** N	
42. Did the older man pull a gun from his belt?	Y **N**	
* Note: Bold Y/N denotes correct answer		

APPENDIX 2: Lawyerese Interview 1

Lawyerese	Response (Y/N)		Conf
1. Were three red cars parked opposite the bus stop?	**Y**	N	
2. Did the younger man have gloves on?	Y	**N**	
3. You've seen that the older man was shaven?	**Y**	N	
4. And it's true the driver wore glasses?	**Y**	N	
5. The older man, he had a watch on his wrist, didn't he?	**Y**	N	
6. The men spoke to each other during the attack, didn't they?	Y	**N**	
7. The car driver left the engine running during the attack didn't he?	**Y**	N	
8. The attacker's car turned out of the junction almost opposite didn't it?	Y	**N**	
9. You've seen that the getaway car was a saloon?	**Y**	N	
10. And the victim was pushed into the back of that car wasn't she?	**Y**	N	
11. Would you say that two women looked on at the attack?	**Y**	N	
12. Did the younger man put a cloth over the victim's mouth?	**Y**	N	
13. When the attacker's car left, did a red car follow?	Y	**N**	
14. You've seen the younger man was tall?	**Y**	N	
15. And would you say he was clean shaved?	**Y**	N	
16. A bus arrived up the road shortly before the attack, didn't it?	**Y**	N	
17. And you've heard that the victim shouted out?	**Y**	N	
18. The older man had a ring on his finger, didn't he?	**Y**	N	
19. The attacker's car had four doors, didn't it?	Y	**N**	
20. You've seen the younger man's hair was receding?	Y	**N**	
21. It's true isn't it that the victim's bag fell to the ground?	**Y**	N	
22. And you've seen the man with the buggy watching from across the road?	Y	**N**	
23. Would you say that the victim struggled?	**Y**	N	
24. Another woman posted a letter into the yellow box, didn't she?	Y	**N**	
25. The woman who walked past the bus stop had two bags, didn't she?	Y	**N**	
26. And you've seen that both men had dark trousers on?	Y	**N**	
27. The car the attacker's arrived in was silver, wasn't it?	**Y**	N	
28. During the attack, the victim fell to the ground, didn't she?	Y	**N**	
29. Would you say the other woman who checked the timetable had short hair?	Y	**N**	

30. Did the younger man have a moustache?	Y	**N**	
31. Would you say both men were wearing jeans?	Y	**N**	
32. You've seen the gun's long barrel?	Y	**N**	
33. The younger man, he wore glasses didn't he?	Y	**N**	
34. And you've seen his light green shirt?	Y	**N**	
35. And the older man's black coat?	**Y**	N	
36. Did the other woman say the time was a quarter-to one?	**Y**	N	
37. Did the victim walk toward the car when it pulled up?	Y	**N**	
38. Did a mirror fall to the ground during the attack?	**Y**	N	
39. You've seen the gang consisted of two men?	Y	**N**	
40. Did the older man have baseball shoes on?	Y	**N**	
41. And you've heard the older man speak to the victim?	**Y**	N	
42. And pull a gun from his belt?	Y	**N**	
*Note: Bold Y/N denotes correct answer			

APPENDIX 3: Negative Feedback Interview 1

Negative Feedback	Response (Y/N)	Conf
1. Were three red cars parked opposite the bus stop?	**Y** N	
2. Did the younger man have gloves on?	Y **N**	
3. You've seen that the older man was shaven? *Think again, the older man was he shaven?*	**Y** N ***Y*** *N*	--------
4. And it's true the driver wore glasses? *Think about that answer, was the driver wearing glasses?*	**Y** N ***Y*** *N*	--------
5. The older man, he had a watch on his wrist, didn't he? *Consider your answer carefully, did he have a watch on?*	**Y** N ***Y*** *N*	--------
6. The men spoke to each other during the attack, didn't they? *Try and be more accurate, did the men speak to each other?*	Y **N** *Y* ***N***	--------
7. The car driver left the engine running during the attack didn't he? *Think back to the event, was the engine left running?*	**Y** N ***Y*** *N*	--------
8. The attacker's car turned out of the junction almost opposite didn't it? *Think again, did the car turn out of the junction opposite?*	Y **N** *Y* ***N***	--------
9. You've seen that the getaway car was a saloon? *Try and be more accurate, the car used was it a saloon?*	**Y** N ***Y*** *N*	--------
10. And the victim was pushed into the back of that car, wasn't she? *Try harder to remember, was the victim pushed into the back of that car?*	**Y** N ***Y*** *N*	--------
11. Would you say that two women looked on at the attack?	**Y** N	

12. Did the younger man put a cloth over the victim's mouth?	**Y** N	
13. When the attacker's car left, did a red car follow?	Y **N**	
14. You've seen the younger man was tall? *Now think back to the event, was the young man tall?*	**Y** N ***Y*** *N*	---
15. And would you say he was clean shaved?	**Y** N	
16. A bus arrived up the road shortly before the attack, didn't it? *Try to be more accurate, did a bus arrive up the road shortly before the attack?*	**Y** N ***Y*** *N*	---
17. And you've heard that the victim shouted out? *Think again, did the victim shout out?*	**Y** N ***Y*** *N*	---
18. The older man had a ring on his finger, didn't he? *Think carefully about your answer, did he have a ring on?*	**Y** N ***Y*** *N*	---
19. The attacker's car had four doors, didn't it? *Try and be more accurate, did the car have four doors?*	Y **N** *Y* ***N***	---
20. You've seen the younger man's hair was receding? *It's important you are correct, was the young man's hair receding?*	Y **N** *Y* ***N***	---
21. It's true isn't it that the victim's bag fell to the ground? *Think back to the event, did her bag fall to the ground?*	**Y** N ***Y*** *N*	---
22. And you've seen the man with the buggy watching from across the road? *Think again, did the man with the buggy watch?*	Y **N** *Y* ***N***	---
23. Would you say that the victim struggled?	**Y** N	
24. Another woman posted a letter into the yellow box, didn't she? *Try harder to remember, did the woman post a letter into the yellow box?*	Y **N** *Y* ***N***	---

25. The woman who walked past the bus stop had two bags, didn't she? *Try to be more accurate, did the woman who walked past the bus stop have two bags?*	Y **N** *Y N*	---
26. And you've seen both men had dark trousers on? *Think again, did both men have dark trousers on?*	Y **N** *Y N*	---
27. The car the attacker's arrived in was silver, wasn't it? *Think carefully about your answer, was the car silver?*	**Y** N ***Y*** *N*	---
28. During the attack, the victim fell to the ground, didn't she? *Think back to the event, did the victim fall to the ground?*	Y **N** *Y N*	---
29. Would you say the other woman who checked the timetable had short hair?	Y **N**	
30. Did the younger man have a moustache?	Y **N**	
31. Would you say both men were wearing jeans?	Y **N**	
32. You've seen the gun's long barrel? *Try to be more accurate, was the gun's barrel long?*	Y **N** *Y N*	---
33. The younger man, he wore glasses didn't he? *Try harder to remember, the younger man did he have glasses on?*	Y **N** *Y N*	---
34. And you've seen his light green shirt? *It's important you are correct, was his shirt light green?*	Y **N** *Y N*	---
35. And the older man's black coat? *Think again, the older man's coat was it black?*	**Y** N ***Y*** *N*	---
36. Did the other woman say the time was a quarter-to one?	**Y** N	
37. Did the victim walk toward the car when it pulled up?	Y **N**	
38. Did a mirror fall to the ground during the attack?	**Y** N	

39. You've seen the gang consisted of two men? *Think carefully about your answer, did the gang consist of two men?*	Y **N** *Y* ***N***	---
40. Did the older man have baseball shoes on?	Y **N**	
41. And you've heard the older man speak to the victim? *Try to be more accurate, did the older man speak to her?*	**Y** N ***Y*** *N*	---
42. And pull a gun from his belt? *Think again, did he pull a gun from his belt?*	Y **N** *Y* ***N***	---
*Note: Bold Y/N denotes correct answer		

APPENDIX 4: Control Interview 2

Control	Response (Y/N)		Conf
1. When the attack occurred was it daylight?	**Y**	N	
2. Would you say that the weather was fine?	**Y**	N	
3. Did two men carry out the attack on the victim?	**Y**	N	
4. Did the victim's bag fall to the ground during the attack?	**Y**	N	
5. Would you say that the older man's coat was black?	**Y**	N	
6. Did the victim shout out during the attack?	**Y**	N	
7. Would you say that the victim struggled?	**Y**	N	
8. Did the older man speak to the victim?	**Y**	N	
9. Did the victim almost fall to the ground during the attack?	**Y**	N	
10. Did the younger man put a cloth over the victim's mouth?	**Y**	N	
11. Were three red cars parked opposite the bus stop?	**Y**	N	
12. And would you say the younger man was tall?	**Y**	N	
13. Did a mirror fall to the ground during the attack?	**Y**	N	
14. Did another woman post a letter into the yellow box?	Y	**N**	
15. Did a bus arrive up the road shortly before the attack?	**Y**	N	
16. Would you say the younger of the men did not have a moustache?	**Y**	N	
17. Did the woman who walked past the bus stop have two bags?	Y	**N**	
18. And the older man was he clean shaved?	**Y**	N	
19. Would you say the men did not speak to each other during the attack?	**Y**	N	

20. Did the other woman who checked the timetable have short hair?	Y	**N**	
21. And did another woman say that the time was a quarter-to one?	**Y**	N	
22. Was the attacker's car silver?	**Y**	N	
23. And would you say that this car had four doors?	Y	**N**	
24. Did this car turn out of the junction almost opposite?	Y	**N**	
25. And the younger man was he clean shaved?	**Y**	N	
26. Was a gun used in the attack?	**Y**	N	
27. When the attacker's car left did a red car follow?	Y	**N**	
28. Would you say the younger of the men was not wearing glasses?	**Y**	N	
29. Was the younger man wearing baseball shoes?	**Y**	N	
30. Did the victim walk toward the car as it pulled up?	Y	**N**	
31. Did the victim have long hair?	**Y**	N	
32. Would you say the older man was going bald?	**Y**	N	
33. Were both of the men wearing dark trousers?	Y	**N**	
34. Would you say both men's trousers were jeans?	Y	**N**	
35. Would you say neither of the men hit the victim during the attack?	**Y**	N	
36. And did the older man pull the gun from his belt?	Y	**N**	
37. Did the older man show the victim an identity card?	**Y**	N	
38. Did the driver who waited in the car wear glasses?	**Y**	N	
*Note: Bold Y/N denotes correct answer			

APPENDIX 5: Lawyerese Interview 2

Lawyerese	Response (Y/N)		Conf
1. Now, it would be fair to say that when the attack occurred it was daylight?	**Y**	N	
2. And that to your best recollection the weather was fine?	**Y**	N	
3. Do you also remember that two men carried out the attack on the victim?	**Y**	N	
4. And that during the attack the victim's bag fell to the ground?	**Y**	N	
5. And you would agree that the older man's coat was black?	**Y**	N	
6. And do you recall that the victim shouted out during the attack?	**Y**	N	
7. Do you recall that the victim struggled?	**Y**	N	
8. And it would be fair to say also that the older man spoke to the victim?	**Y**	N	
9. And you would agree that during the attack the victim did not fall to the ground completely?	**Y**	N	
10. And do you remember the younger man putting a cloth over the victim's mouth?	**Y**	N	
11. Can you remember three red cars parked opposite the bus stop?	**Y**	N	
12. To your best recollection, the younger man was tall, wasn't he?	**Y**	N	
13. And you remember a mirror fell to the ground during the attack?	**Y**	N	
14. And it's true isn't it that another woman posted a letter into the yellow box?	Y	**N**	
15. And do you remember that a bus arrived up the road shortly before the attack?	**Y**	N	

16. Do you accept the younger man did not have a moustache?	**Y**	N	
17. Do you recall that the woman who walked past the bus stop had two bags?	Y	**N**	
18. And do you also remember that the older man was clean shaved?	**Y**	N	
19. You would agree ... that the men did not speak to each other during the attack?	**Y**	N	
20. The other woman who checked the timetable she had short hair, didn't she?	Y	**N**	
21. And you do recall that another woman said that the time was a quarter-to-one?	**Y**	N	
22. Can you remember the attacker's car was silver?	**Y**	N	
23. Wouldn't it also be correct to say that this car had four doors?	Y	**N**	
24. And do you remember that car turned out of the junction almost opposite prior to the attack?	Y	**N**	
25. And do you recall that the younger man was also clean shaved?	**Y**	N	
26. You would agree that a gun was used in the attack?	**Y**	N	
27. And do you remember that when the attacker's car left a red car followed?	Y	**N**	
28. You would accept that the younger of the men was not wearing glasses, wouldn't you?	**Y**	N	
29. It would be fair to say wouldn't it that the younger of the men was wearing baseball shoes?	**Y**	N	
30. Its true that the victim walked toward the car as it pulled up isn't it?	Y	**N**	
31. And you would agree that the victim's hair was long?	**Y**	N	

32. Do you recall that the older man was going bald?	**Y**	N	
33. And it's right isn't it that you've seen both men were wearing dark trousers?	Y	**N**	
34. And do you remember these trousers were jeans?	Y	**N**	
35. You would agree though that during the attack neither of the men hit the victim?	**Y**	N	
36. And you remember the older man pulled the gun from his belt?	Y	**N**	
37. Prior to the attack can you remember the older man showing the victim an identity card?	**Y**	N	
38. And to your best recollection do you recall the driver who waited in the car was wearing glasses?	**Y**	N	
*Note: Bold Y/N denotes correct answer			

APPENDIX 6: Lawyerese with Negative Feedback Interview 2

Lawyerese with Negative Feedback	Response (Y/N)		Conf
1. Now, it would be fair to say that when the attack occurred it was daylight? *Is it possible that it was daylight?*	**Y**	N	
2. And that to your best recollection the weather was fine? *Do you accept that the weather could have possibly been fine?*	**Y**	N	
3. Do you also remember that two men carried out the attack on the victim? *Is it possible that you're mistaken and that two men were the perpetrators?*	**Y**	N	
4. And that during the attack the victim's bag fell to the ground? *Do you accept it's possible the victim's bag did fall to the ground?*	**Y**	N	
5. And you would agree that the older man's coat was black? *Is it possible you're mistaken and his coat was actually black?*	**Y**	N	
6. And do you recall that the victim shouted out during the attack? *Would you accept the possibility that she did shout out?*	**Y**	N	
7. Do you recall that the victim struggled? *Is it possible that she was struggling and you are mistaken about that?*	**Y**	N	
8. And it would be fair to say also that the older man spoke to the victim? *Is it possible that the older man did speak to the victim?*	**Y**	N	
9. And you would agree that during the attack the victim did not fall to the ground completely? *Is it possible she didn't fall to the ground completely?*	**Y**	N	
10. And do you remember the younger man putting a cloth over the victim's mouth? *Would you accept that he could have put a cloth over her mouth and you're mistaken?*	**Y**	N	

11. Can you remember three red cars parked opposite the bus stop? *Do you accept it is possible that there were three red cars parked opposite the bus stop?*	**Y**	N	
12. To your best recollection, the younger man was tall, wasn't he? *It's possible he was tall though and you're mistaken?*	**Y**	N	
13. And you remember a mirror fell to the ground during the attack? *Is it possible that a mirror did fall to the ground at that time?*	**Y**	N	
14. And it's true isn't it that another woman posted a letter into the yellow box? *Is it possible that she posted the letter into the yellow box?*	Y	**N**	
15. And do you remember that a bus arrived up the road shortly before the attack? *Is it possible that you're mistaken and a bus did arrive up the road?*	**Y**	N	
16. Do you accept the younger man did not have a moustache? *Do you accept it's possible that the younger man may have had a moustache?*	**Y**	N	
17. Do you recall that the woman who walked past the bus stop had two bags? *Is it at all possible that she had two bags?*	Y	**N**	
18. And do you also remember that the older man was clean shaved? *Is it possible he was actually clean shaved?*	**Y**	N	
19. You would agree ... that the men did not speak to each other during the attack? *Do you accept that it is possible they didn't speak to each other and you're mistaken?*	**Y**	N	
20. The other woman who checked the timetable she had short hair, didn't she? *Do you accept the possibility that her hair could have been short?*	Y	**N**	
21. And you do recall that another woman said that the time was a quarter-to-one?	**Y**	N	

But it's possible that she might have said the time was a quarter-to one?			
22. Can you remember the attacker's car was silver? *Is it possible it was silver?*	**Y**	N	
23. Wouldn't it also be correct to say that this car had four doors? *Isn't it possible it had four doors and you're mistaken?*	Y	**N**	
24. And do you remember that car turned out of the junction almost opposite prior to the attack? *It is possible that the car turned out of that junction though?*	Y	**N**	
25. And do you recall that the younger man was also clean shaved? *Is it a possibility that he was clean shaved?*	**Y**	N	
26. You would agree that a gun was used in the attack? *Do you accept it is possible that a gun was used?*	**Y**	N	
27. And do you remember that when the attacker's car left a red car followed? *Do you accept it's possible that a red car followed when the attacker's car left?*	Y	**N**	
28. You would accept that the younger of the men was not wearing glasses, wouldn't you? *Is it possible you're mistaken and he wasn't wearing glasses?*	**Y**	N	
29. It would be fair to say wouldn't it that the younger of the men was wearing baseball shoes? *Do you accept it's possible you may be mistaken and he was wearing baseball shoes?*	**Y**	N	
30. Its true that the victim walked toward the car as it pulled up isn't it? *It is possible she walked toward the car though?*	Y	**N**	
31. And you would agree that the victim's hair was long? *Do you accept her hair could have possibly been long?*	**Y**	N	
32. Do you recall that the older man was going bald? *It's possible he was going bald though?*	**Y**	N	

33. And it's right isn't it that you've seen both men were wearing dark trousers? *You could be mistaken and it's possible then that both trousers were dark?*	Y	**N**	
34. And do you remember these trousers were jeans? *But it's possible these trousers were actually jeans?*	Y	**N**	
35. You would agree though that during the attack neither of the men hit the victim? *Do you accept that it is possible neither of the men hit the victim?*	**Y**	N	
36. And you remember the older man pulled the gun from his belt? *Is it possible he pulled the gun from his belt?*	Y	**N**	
37. Prior to the attack can you remember the older man showing the victim an identity card? *Do you accept it's possible he showed her an identity card?*	**Y**	N	
38. And to your best recollection do you recall the driver who waited in the car was wearing glasses? *Is it possible the driver had glasses on?*	**Y**	N	

*Note: Bold Y/N denotes correct answer

APPENDIX 7: Questionnaire (including Fear of Negative Evaluation Scale and Social Distress Scale; Part 1, Discomfort Scale; Part 2, and Self-rating Scales)

QUESTIONNAIRE

PART 1:

Read each item carefully and decide the degree to which the statement is characteristic or true of you. Then place a number between "1" and "5" in the space according to the following scale.

1 = The statement is **NOT** characteristic of me.
2 = The statement is **SLIGHTLY** characteristic of me.
3 = The statement is **MODERATELY** characteristic of me.
4 = The statement is **VERY** characteristic of me.
5 = The statement is **EXTREMELY** characteristic of me.

_____ **1.** I am unconcerned even if I know people are forming an unfavourable impression of me
_____ **2.** I am not afraid that people will find fault with me
_____ **3.** I often find social occasions upsetting
_____ **4.** I usually feel calm and comfortable at social occasions
_____ **5.** When I am talking to someone I do not worry about what they might be thinking about me
_____ **6.** I often feel nervous or tense in casual get-togethers in which both sexes are present
_____ **7.** I am usually nervous with people unless I know them well
_____ **8.** I am usually worried about what kind of impression I make
_____ **9.** I usually feel uncomfortable when I am in a group of people I don't know
_____ **10.** I usually feel relaxed when I meet someone for the first time
_____ **11.** If I know someone is judging me, it has little effect on me
_____ **12.** I often feel on edge when I am with a group of people
_____ **13.** I often worry that I will say or do the wrong things
_____ **14.** I find it easy to relax with other people
_____ **15.** I worry about what people will think of me even when I know it doesn't matter
_____ **16.** I feel relaxed even in unfamiliar social situations
_____ **17.** I am frequently afraid of other people noticing my shortcomings
_____ **18.** I rarely worry about what kind of impression I am making on someone
_____ **19.** It is easy for me to relax when I am with strangers
_____ **20.** I am afraid others will not approve of me
_____ **21.** Other people's opinions do not bother me
_____ **22.** I am usually at ease when talking to someone of the opposite

sex

_____ **23.** I usually feel relaxed when I am with a group of people

_____ **24.** Being introduced to people makes me tense and nervous

_____ **25.** Sometimes I think I am too concerned with what other people think of me

_____ **26.** I am seldom at ease in a large group of people

PART 2:

Again read each item carefully and decide the degree to which the statement was characteristic or true of you in this situation. Then place a number between "1" and "5" in the space according to the following scale.

1 = The statement is **NOT** characteristic of me.
2 = The statement is **SLIGHTLY** characteristic of me.
3 = The statement is **MODERATELY** characteristic of me.
4 = The statement is **VERY** characteristic of me.
5 = The statement is **EXTREMELY** characteristic of me.

_____ **1.** I felt calm and able to answer questions comfortably

_____ **2.** At times I felt under pressure

_____ **3.** I was often not confident about the answers to the questions

_____ **4.** I did not feel tense or nervous about answering the questions

_____ **5.** I was relieved when the interview had finished

_____ **6.** I was confident in my memory ability

PART 3:

On the following scale please circle the number that most accurately describes the way you felt/feel about yourself.

1. How accurate do you think you were?

1	2	3	4	5
Extremely Inaccurate	**A little Inaccurate**	**Moderately Accurate**	**Very Accurate**	**Extremely Accurate**

2. How confident were you in your performance?

1	2	3	4	5
Not at all Confident	**A little Confident**	**Moderately Confident**	**Very Confident**	**Extremely Confident**

3. How good a witness do you think you were?

1	2	3	4	5
Extremely Poor	**Slightly Poor**	**Moderately Good**	**Very Good**	**Extremely Good**

4. How happy would you be to give evidence in court?

1	2	3	4	5
Not at all Happy	**A little Unhappy**	**Moderately Happy**	**Very Happy**	**Extremely Happy**

5. How well do you think you would cope with giving evidence in court?

1	2	3	4	5
Not at all Well	**A little Well**	**Moderately Well**	**Very Well**	**Extremely Well**

6. How good a witness do you think you would be in court?

1	2	3	4	5
Extremely Poor	**Slightly Poor**	**Moderately Good**	**Very Good**	**Extremely Good**

Thank you for your co-operation in this study. Your time and efforts are very much appreciated.

APPENDIX 8: Control Interview 3

Control	Response (Y/N)		Conf
1. When the attack occurred was it daylight?	**Y**	N	
2. Would you say that the weather was fine?	**Y**	N	
3. Did two men carry out the attack on the victim?	**Y**	N	
4. Did the victim's bag fall to the ground during the attack?	**Y**	N	
5. Would you say that the older man's coat was black?	**Y**	N	
6. Did the victim shout out during the attack?	**Y**	N	
7. Would you say that the victim struggled?	**Y**	N	
8. Did the older man speak to the victim?	**Y**	N	
9. Did the victim almost fall to the ground during the attack?	**Y**	N	
10. Did the younger man put a cloth over the victim's mouth?	**Y**	N	
11. Were three red cars parked opposite the bus stop?	**Y**	N	
12. And would you say the younger man was tall?	**Y**	N	
13. Did a mirror fall to the ground during the attack?	**Y**	N	
14. Did the victim have trousers on?	Y	**N**	
15. Did a bus arrive up the road shortly before the attack?	**Y**	N	
16. Would you say the younger of the men did not have a moustache?	**Y**	N	
17. Did the gun have a long barrel?	Y	**N**	
18. And the older man was he clean shaved?	**Y**	N	
19. Would you say the men did not speak to each other during the attack?	**Y**	N	
20. Did the younger man point the gun at the witness?	Y	**N**	

21. And did another woman say that the time was a quarter-to one?	**Y**	N	
22. Was the attacker's car silver?	**Y**	N	
23. Did the victim have glasses on?	Y	**N**	
24. Did the victim have her hair down?	Y	**N**	
25. And the younger man was he clean shaved?	**Y**	N	
26. Was a gun used in the attack?	**Y**	N	
27. Did the woman who posted the letter have a dress on?	Y	**N**	
28. Would you say the younger of the men was not wearing glasses?	**Y**	N	
29. Was the younger man wearing baseball shoes?	**Y**	N	
30. Did the younger of the men hit the victim during the attack?	Y	**N**	
31. Did the victim have long hair?	**Y**	N	
32. Would you say the older man was going bald?	**Y**	N	
33. When the attacker's car left, did a police car follow?	Y	**N**	
34. Was the younger man's coat black?	Y	**N**	
35. Would you say neither of the men hit the victim during the attack?	**Y**	N	
36. Did the younger man speak to the victim?	Y	**N**	
37. Did the older man show the victim an identity card?	**Y**	N	
38. Did the driver who waited in the car wear glasses?	**Y**	N	

* Note: Bold Y/N denotes correct answer

APPENDIX 9: Lawyerese Interview 3

Lawyerese	Response (Y/N)		Conf
1. Now, it would it be fair to say that when the attack occurred it was daylight?	**Y**	N	
2. And that to your best recollection the weather was fine?	**Y**	N	
3. Do you also remember that two men carried out the attack on the victim?	**Y**	N	
4. And that during the attack the victim's bag fell to the ground?	**Y**	N	
5. And you would agree that the older man's coat was black?	**Y**	N	
6. And do you recall that the victim shouted out during the attack?	**Y**	N	
7. Do you recall that the victim struggled?	**Y**	N	
8. And it would be fair to say also that the older man spoke to the victim?	**Y**	N	
9. And you would agree that during the attack the victim did not fall to the ground completely?	**Y**	N	
10. And do you remember the younger man putting a cloth over the victim's mouth?	**Y**	N	
11. Can you remember three red cars parked opposite the bus stop?	**Y**	N	
12. To your best recollection, the younger man was tall, wasn't he?	**Y**	N	
13. And you remember a mirror fell to the ground during the attack?	**Y**	N	
14. If I said to you the victim was wearing trousers this is something you would agree with isn't it?	Y	**N**	

15. And do you remember that a bus arrived up the road shortly before the attack?	**Y**	N	
16. Do you accept that the younger man did not have a moustache?	**Y**	N	
17. You do remember that the gun had a long barrel?	Y	**N**	
18. And do you also remember the older man was clean shaved?	**Y**	N	
19. Would you agree ... that the men did not speak to each other during the attack?	**Y**	N	
20. The younger man pointed the gun at the witness didn't he?	Y	**N**	
21. And do you recall that another woman said the time was a quarter-to one?	**Y**	N	
22. Can you remember the attacker's car was silver?	**Y**	N	
23. And it's true isn't it that the victim was wearing glasses?	Y	**N**	
24. Do you recall that the victim's hair was down?	Y	**N**	
25. And do you recall that the younger man was also clean shaved?	**Y**	N	
26. You would agree that a gun was used in the attack?	**Y**	N	
27. Do you remember that the woman who posted the letter had a dress on?	Y	**N**	
28. You would accept that the younger of the men was not wearing glasses, wouldn't you?	**Y**	N	
29. It would be fair to say wouldn't it that the younger of the men was wearing baseball shoes?	**Y**	N	

30. And isn't it also true that the younger man hit the victim?	Y	**N**	
31. And would you agree that the victim's hair was long?	**Y**	N	
32. Do you recall that the older man was going bald?	**Y**	N	
33. Isn't it also correct that when the attacker's car left a police car followed?	Y	**N**	
34. And do you remember the younger man's coat was black?	Y	**N**	
35. You would agree though that during the attack neither of the men hit the victim?	**Y**	N	
36. Do you recall the younger man spoke to the victim?	Y	**N**	
37. Prior to the attack can you remember the lder man showing the victim an identity card?	**Y**	N	
38. And to your best recollection do you recall the driver who waited in the car was wearing glasses?	**Y**	N	
* Note: Bold Y/N denotes correct answer			

APPENDIX 10: Lawyerese with Negative Feedback 3

Lawyerese with Negative Feedback	Response (Y/N)	Conf
1. Now, it would it be fair to say that when the attack occurred was it daylight? *Is it possible that it was daylight?*	**Y** N	
2. And that to your best recollection the weather was fine? *Do you accept that the weather could have possibly been fine?*	**Y** N	
3. Do you also remember that two men carried out the attack on the victim? *Is it possible that you're mistaken and that two men were the perpetrators?*	**Y** N	
4. And that during the attack the victim's bag fell to the ground? *Do you accept it's possible the victim's bag did fall to the ground?*	**Y** N	
5. And you would agree that the older man's coat was black? *Is it possible you're mistaken and his coat was actually black?*	**Y** N	
6. And do you recall that the victim shouted out during the attack? *Would you accept the possibility that she did shout out?*	**Y** N	
7. Do you recall that the victim struggled? *Is it possible that she was struggling and you are mistaken about that?*	**Y** N	
8. And it would be fair to say also that the older man spoke to the victim? *Is it possible that the older man did speak to the victim?*	**Y** N	
9. And you would agree that during the attack the victim did not fall to the ground completely? *Is it possible she didn't fall to the ground completely?*	**Y** N	

10. And do you remember the younger man putting a cloth over the victim's mouth? *Would you accept that he could have put a cloth over her mouth and you're mistaken?*	**Y**	N	
11. Can you remember three red cars parked opposite the bus stop? *Do you accept it is possible that there were three red cars parked opposite the bus stop?*	**Y**	N	
12. To your best recollection, the younger man was tall, wasn't he? *It's possible he was tall though and you're mistaken?*	**Y**	N	
13. And you remember a mirror fell to the ground during the attack? *Is it possible that a mirror did fall to the ground at that time?*	**Y**	N	
14. If I said to you the victim was wearing trousers this is something you would agree with isn't it? *Could it be possible that she was actually wearing trousers and that you are mistaken?*	Y	**N**	
15. And do you remember that a bus arrived up the road shortly before the attack? *Is it possible that you're mistaken and a bus did arrive up the road?*	**Y**	N	
16. Do you accept the younger man did not have a moustache? *Do you accept it's possible that the younger man may have had a moustache?*	**Y**	N	
17. It's true isn't it that the gun had a long barrel? *Is it at all possible the gun had a long barrel?*	Y	**N**	
18. And do you also remember that the older man was clean shaved? *Is it possible he was actually clean shaved?*	**Y**	N	
19. You would agree ... that the men did not speak to each other during the attack?	**Y**	N	

Do you accept that it is possible they didn't speak to each other and you're mistaken?			
20. The younger man pointed the gun at the witness didn't he? *Is it possible that the younger man did point the gun at the witness?*	Y	**N**	
21. And you do recall that another woman said that the time was a quarter-to-one? *But it's possible that she might have said the time was a quarter-to one?*	**Y**	N	
22. Can you remember the attacker's car was silver? *Is it possible it was silver?*	**Y**	N	
23. And it's true isn't it that the victim was wearing glasses? *Is it at all possible she was wearing glasses?*	Y	**N**	
24. Do you recall that the victim's hair was down? *Is it possible her hair was actually down?*	Y	**N**	
25. And do you recall that the younger man was also clean shaved? *Is it a possibility that he was clean shaved?*	**Y**	N	
26. You would agree that a gun was used in the attack? *Do you accept it is possible that a gun was used?*	**Y**	N	
27. Do you remember that the woman who posted the letter had a dress on? *Do you accept she may have possibly had a dress on?*	Y	**N**	
28. You would accept that the younger of the men was not wearing glasses, wouldn't you? *Is it possible you're mistaken and he wasn't wearing glasses?*	**Y**	N	
29. It would be fair to say wouldn't it that the younger of the men was wearing baseball shoes? *Do you accept it's possible you may be mistaken*	**Y**	N	

and he was wearing baseball shoes?			
30. And isn't it also true that the younger man hit the victim? *Is it possible that he did hit her?*	Y	**N**	
31. And you would agree that the victim's hair was long? *Do you accept her hair could have possibly been long?*	**Y**	N	
32. Do you accept the older man was going bald? *Is it possible he was going bald?*	**Y**	N	
33. Isn't it also correct that when the attacker's car left a police car followed? *It is possible a police followed the attacker's car when it left though and you are mistaken about this?*	Y	**N**	
34. And do you remember the younger man's coat was black? *Is it possible that it was actually black?*	Y	**N**	
35. You would agree though that during the attack neither of the men hit the victim? *Do you accept that it is possible neither of the men hit the victim?*	**Y**	N	
36. Do you recall the younger man spoke to the victim? *It is possible he spoke to her though?*	Y	**N**	
37. Prior to the attack can you remember the older man showing the victim an identity card? *Do you accept it's possible he showed her an identity card?*	**Y**	N	
38. And to your best recollection do you recall the driver who waited in the car was wearing glasses? *Is it possible the driver had glasses on?*	**Y**	N	
* Note: Bold Y/N denotes correct answer			

APPENDIX 11: Jurors' Perception's Questionnaire

QUESTIONNAIRE

Name	Age	Dept./Study/Occupation	Contact - *email/tel*

Please read the following questions in turn and decide, on the scale following each question, the response that you feel most accurately reflects your experience of the taped interview you have just heard.

1. **How ACCURATE do you think the witness was?**

1	2	3	4	5
Extremely Inaccurate	A little Inaccurate	Moderately Accurate	Very Accurate	Extremely Accurate

2. **How CONFIDENT do you think the witness was?**

1	2	3	4	5
Not at all Confident	A little Confident	Moderately Confident	Very Confident	Extremely Confident

3. **How STRESSED do you think the witness was?**

1	2	3	4	5
Not at all Stressed	A little Stressed	Moderately Stressed	Very Stressed	Extremely Stressed

4. **How GOOD a witness do you think the witness actually was?**

1	2	3	4	5
Extremely Poor	Slightly Poor	Moderately Good	Very Good	Extremely Good

5. **How FAIR do you think the questioning was?**

1	2	3	4	5
Not at all Fair	A little Fair	Moderately Fair	Very Fair	Extremely Fair

Appendix 12 Correlations (*r*) between actual performance variables, witnesses' own performance ratings, distress scale, fear of negative evaluation scale, and discomfort scale (n=60).

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1discmft	-	.28*	.25	.00	-.04	.21	-.19	.03	.00	-.02	.06	-.13	-.08	-.04	-.18	-.27*	-44***	-.31*	-.16	-.13	-.20
2fne		-	.54***	-.04	.09	.11	-.15	.12	-.04	.12	.08	.17	-.12	.26*	.04	-.08	-.12	-.07	-.29*	-.25	-.26*
3distress			-	.00	.18	.14	-.13	-.03	-.20	.00	.10	.05	-.01	.20	-.00	-.21	-.30*	-.25	-.25	-.07	-.16
4caall				-	.30*	.53***	.36**	-.15	.05	-.24	-.00	.09	.06	.04	-.07	-.00	-.02	.10	-.01	.08	.04
5caeasy					-	.10	-.10	-.22	-45***	-.06	.07	.04	.18	.20	-.04	-.05	-.07	-.09	-.10	-.13	-.15
6camod						-	-42**	-.17	.11	-.40**	.10	.12	-.02	.19	-.03	-.16	.01	.16	.15	.21	.10
7cadiff							-	.19	-.02	.41	-.15	.10	.00	-.01	.15	.16	-.00	-.04	-.10	-.08	.02
8nocorall								-	.50***	.62***	.44***	.35**	-.04	.35**	.22	.32*	.00	.10	-.09	-.09	.01
9ncresy									-	.09	-.10	.27*	-.09	.04	.27*	.23	.26*	.33*	.13	.18	.22
10ncrmod										-	-.24	.24	.04	.34**	.15	.21	.06	-.04	-.15	-.17	-.10
11nocrdif											-	.06	-.04	.13	-.05	.09	-.26*	-.04	-.06	-.09	-.04
12xconfall												-	-.08	.86***	.80***	.28*	.22	.01	.12	.06	-.02
13xcoesy													-	-.11	-.08	.20	.01	.22	.13	-.02	.19
14xcomod														-	.52***	.17	.09	-.04	.03	-.02	-.13
15xcodiff															-	.16	.26*	-.05	.14	.09	.07
16accy																-	.53***	.47***	.08	.26*	.35**
17gdwtns																	-	.42***	.28*	.32*	.37
18perfco																		-	.06	.20	.23
19evidcrt																			-	.33*	.53***
20copecrt																				-	.62***
21gdwitcrt																					-

Note: *$p<.05$, **$p<.01$, ***$p<.001$ (2-tailed). Items 1-3 are scales. Items 4-15 are performance measures. Items 16-21 are witnesses' own ratings. See Appendix 20 for Key.

Appendix 13 Correlations (*r*) between actual performance variables, witnesses' own performance ratings, distress scale, fear of negative evaluation scale, and discomfort scale (n=60).

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1discmft	-	.28*	.25	.00	-.19	-.07	.01	-.15	-.23	-.13	-.18	-.09	-.14	-.27*	-44***	-.31*	-.16	-.13	-.20
2fne		-	.54***	-.04	-.15	.05	.03	.11	-.03	.17	.04	.07	-.04	-.08	-.12	-.07	-.29*	-.25	-.26*
3distress			-	.00	-.13	.11	.02	-.01	-.02	.05	-.01	-.08	-.03	-.21	-.30*	-.25	-.25	-.07	-.16
4caall				-	.36**	-.21	-.14	.16	.13	.09	-.07	.26*	.17	-.00	-.02	.10	-.01	.08	.04
5cadiff					-	-.14	-.28*	.25	.41***	.10	.15	.30*	.42***	.16	-.00	-.04	-.10	-.08	.02
6flsalmall						-	.60***	.10	.02	.46***	.50***	-51***	-.35**	-.17	.05	-.17	.08	.05	-.15
7flsalmdif							-	.17	-.33**	.35**	.51***	-.21	-81***	-.12	.25	-.01	.10	.24	.03
8hitsall								-	.57***	.79***	.62***	.81***	.25	.31*	.21	.14	.05	.12	.07
9hitsdiff									-	.44***	.51***	.48***	.82***	.19	.00	-.03	-.01	-.10	-.01
10xconall										-	.80***	.41***	.07	.28*	.22	.01	.12	.06	-.02
11xcondif											-	.24	.01	.16	.26*	-.05	.14	.09	.07
12%accall												-	.43***	.36**	.15	.22	-.00	.08	.15
13%accdif													-	.19	-.15	-.01	-.07	-.21	-.03
14accy														-	.53***	.47***	.08	.26*	.35**
15gdwtns															-	.42***	.28*	.32*	.37**
16perfco																-	.06	.20	.23
17evidcrt																	-	.33*	.53***
18copecrt																		-	.62***
19gdwitcrt																			-

Note: *$p<.05$, **$p<.01$, ***$p<.001$ (2-tailed). Items 1-3 are scales. Items 4-13 are performance measures. Items 14-19 are witnesses' own ratings. See Appendix 20 for Key.

APPENDIX 14: Scoring Rules for Clustering and Switching (FAS)

For each protocol, three scores were calculated, including the total number of words generated, mean cluster size, and number of switches.

Total number of correct words generated. This was calculated as the sum of all words produced, excluding errors and repetitions.

Mean cluster size. Cluster size was counted starting with the second word in a cluster. That is, a single word was given a cluster size of 0, two words had a cluster size of 1, three words had a cluster size of 2, and so on. The mean cluster size was computed for the three phonemic trials.

Number of switches. This was calculated as the total number of transitions between clusters, including single words, for the three phonemic trials combined.

Phonemic Fluency: Clusters on phonemic fluency trials consisted of successively generated words that shared any of the following phonemic characteristics:

First letters: words beginning with same first two letters, such as *arm* and *art*.

Rhymes: words that rhyme, such as *sand* and *stand*.

First and last sounds: words differing only by a vowel sound, regardless of the actual spelling, such as *sat, seat, soot, sight,* and *sought*.

Homonyms: Words with two or more different spellings, such as *some* and *sum*.

Appendix 15 Correlations (*r*) between actual performance variables, discomfort, and neurological tests (FAS); (n=21).

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1acc5	-	.25	.12	.20	-.08	.51*	-.55*	.24	.38	.12	-.07	.26	-.50*	-.21	-.09	-.01	.12
2acc6		-	.40	.52*	.07	-.30	-.14	-.24	-.14	-.18	.40	.47*	-.26	.01	.48**	.43	-.13
3conf5			-	.78***	.12	.34	.09	.13	.01	.41	.13	.42	-.27	-.34	.41	.33	-.33
4conf6				-	.01	.07	-.02	.12	.03	.21	.08	.19	-.22	-.48*	.52**	.41	-.43
5caall5					-	.21	.03	-.08	.50*	.07	.50*	.34	.12	-.04	.11	.15	.14
6caall6						-	-.13	.34	.50*	.63**	-.25	.28	-19	-.36	-.23	-.18	.08
7caeasy5							-	-.02	-.43	.20	.14	-.33	.29	-.08	.36	.40	.12
8caeasy6								-	.25	.03	-.41	.13	-.09	-.26	-.08	-.04	-.10
9camod5									-	.34	-.17	.17	-.03	-.17	-.14	-.08	.05
10camod6										-	-.13	.14	.04	-.31	.17	.14	-.22
11cadiff5											-	.35	.13	.15	.41*	.36	-.09
12cadiff6												-	.01	.13	.16	.10	-.19
13discft5													-	.55**	.09	.12	.12
14discft6														-	-.12	-.03	.35
15switches															-	.95***	-.11
16totwfas																-	.16
17xclusfas																	-

Note: *$p<.05$, **$p<.01$, ***$p<.001$ (2-tailed). Items 1-12; performance measures. Items 13-14; scale. Items 15-17; neurological measures. Items 15x2; 15x4; 15x11; (1-tailed). See Appendix 20 for Key.

Appendix 16 Correlations (*r*) between actual performance variables, discomfort, and neurological tests (FAS); (n=21).

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1acceasy5	-	.41	.38	.09	-.21	-.41	.28	.20	.02	.04	-.10	.05	-.44*	-.10	-.36	-.40	-.07
2acceasy6		-	.47*	.27	-.45*	-.43	.37	.50*	.42	.31	.39	.53*	-.15	-.05	.28	.13	-.29
3accmod5			-	.54*	-.42	-.49*	-.04	.02	.21	.21	.22	.37	-.28	.20	.25	.20	-.11
4accmod6				-	-.20	-69***	.32	.31	.30	.52*	.17	.43	-.17	.06	.33	.25	-.34
5accdiff5					-	.49*	.03	.08	.01	.13	-.12	-.16	-.13	-.33	-.11	.02	.25
6accdiff6						-	-.37	-.28	-.22	-.35	-.17	-.38	.04	-.02	-.09	.08	.54*
7coneasy5							-	.81***	.72***	.49*	.50*	.53*	-.21	-.26	.25	.19	-.23
8coneasy6								-	.80***	.74***	.61**	.79***	-.34	-.50*	.35	.26	-.30
9conmd5									-	.63**	.80***	.74***	-.34	-.41	.39	.31	-.34
10conmd6										-	.46*	.85***	-.12	-.45*	.51*	.42	-.44*
11condif5											-	.66**	-.14	-.24	.50*	.42	-.29
12condif6												-	-.14	-.37	.62**	.48*	-.46*
13discft5													-	.55**	.09	.12	.12
14discft6														-	-.12	-.03	.35
17switches															-	.95***	-.11
18totwfas																-	.16
19xclusfas																	-

Note: *$p<.05$, **$p<.01$, ***$p<.001$ (2-tailed). Items 1-12; performance measures. Items 13-14; scale. Items 15-17; neurological measures. See Appendix 20 for Key.

Appendix 17 Correlations (*r*) between actual performance variables, discomfort, and neurological tests (FAS); (n=21).

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1cocoall5	-	.75***	.80***	.73***	.79***	.52*	.14	.69***	.45*	.52*	.48*	.37	-.03	-.25	.34	.30	-.19
2cocoall6		-	.63**	.91***	.80***	.90***	.12	.71***	.52*	.80***	.35	.42	-.24	-.57**	.40	.30	-.41
3cocoesy5			-	.71***	.77***	.38	.07	.62**	.57**	.47*	.24	.31	-.14	-.35	.36	.29	-.30
4cocoesy6				-	.77***	.71***	.01	.66**	.54*	.71***	.26	.42	-.23	-.61**	.25	.19	-.29
5ccomod5					-	.67**	.02	.73***	.68**	.50*	.47*	.27	-.25	-.45*	.30	.24	-.30
6ccomod6						-	.14	.54*	.36	.70***	.43	.36	-.10	-.53*	.43*	.37	-.33
7cocodif5							-	.43	-.08	-.02	.36	-.01	.20	-.15	.31	.33	-.03
8cocodif6								-	.63	.52*	.37	.47*	-.29	-.52*	.42	.35	-.23
9coinall5									-	.42	.32	.35	-.52*	-.40	.25	.16	.35
10coinall6										-	.28	.48*	-.10	-.30	.58**	.45*	-.46*
11cinesy5											-	.30	.15	-.11	.53*	.57	.05
12cinesy6												-	-.27	-.09	.26	.29	.12
13discft5													-	.55**	.09	.12	.12
14discft6														-	-.12	-.03	.35
15switches															-	.95***	-.11
16totwfas																-	.16
17xclusfas																	-

Note: **p*<.05, ***p*<.01, ****p*<.001 (2-tailed). Items 1-12; performance measures. Items 13-14; scale. Items 15-17; neurological measures. See Appendix 20 for Key.

Appendix 18 Correlations (*r*) between actual performance variables, witnesses' self-reports, discomfort scale, and neurological tests (FAS); (n=21).

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1coinmd5	-	.20	.61**	.48*	.41	.19	.19	.48*	.46*	.51*	.40	.37	-.28	-.38	.25	.17	-.21
2coinmd6		-	.01	.34	.07	-.16	.20	.33	.27	.19	.04	.51*	.09	-.37	.21	.16	-.25
3coindif5			-	.28	.54*	.34	.11	.41	.56**	.59**	.31	.18	-.22	-.29	.04	-.03	-.36
4coindif6				-	-.16	.29	-.48*	.24	.10	.27	.06	.35	-.20	-.33	.28	.21	-.22
5sracc5					-	.29	.46*	.36	.73***	.55**	.43	.38	-.16	-.15	.29	.33	.06
6sracc6						-	.07	.32	.51*	.47*	.44*	.47*	-.46*	-.20	.10	.02	-.27
7srconp5							-	.12	.51*	.11	.35	.20	-.22	-.10	-.02	.01	.05
8srconp6								-	.54*	.83***	.63**	.74***	-.33	-70***	.25	.11	-.51*
9srgdwt5									-	.64**	.65**	.56**	-.54*	-.56**	-.01	-.02	-.16
10srgdwt6										-	.47*	.70***	-.39	-.54*	.34	.24	-.41
11srcopct5											-	.56**	-.48*	-.48*	-.12	-.24	-.42
12srcopct6												-	-.53*	-.58**	.32	.21	-.37
13discft5													-	.55**	.09	.12	.12
14discft6														-	-.12	-.03	.35
15switches															-	.95***	-.11
16totwfas																-	.16
17xclusfas																	-

Note: *$p<.05$, **$p<.01$, ***$p<.001$ (2-tailed). Items 1-4; performance measures. Items 5-12; self-reports. Items13-14; scale. Items 15-17; neurological measures. See Appendix 20 for Key.

Appendix 19 Correlations (*r*) between witnesses' self-reports, discomfort scale, and neurological tests (FAS); (n=21).

Measure	1	2	3	4	5	6	7	8	9
1srevdct5	-	.72***	.38	.57**	-.57**	-.35	.10	.02	-.17
2srevdct6		-	.44*	.80***	-.57**	-.71***	.39	.33	-.12
3srgdwtct5			-	.47*	-.39	-.24	.18	.16	-.07
4srgdwtct6				-	-.70***	-.61**	.12	.04	-.25
5discft5					-	.55**	.09	.12	.12
6discft6						-	-.12	-.03	.35
7switches							-	.95***	-.11
8totwfas								-	.16
9xclusfas									-

Note: *p<.05, **p<.01, ***p<.001 (2-tailed). Items 1-4; self-reports. Items 5-6; scale. Items 7-9; neurological measures. See Appendix 20 for Key.

APPENDIX 20: Key to measures contained in Appendices 12-13 and 15-19

Appendices 12-13 and 15-19

discmft	Scale; measure of discomfort at retrieval
fne	Scale; measure of fear of negative evaluation
distress	Scale; measure of social distress
caall	Within-subjects confidence-accuracy correlation for overall interview
caeasy	Within-subjects confidence-accuracy correlation for easy items
camod	Within-subjects confidence-accuracy correlation for moderate items
cadiff	Within-subjects confidence-accuracy correlation for difficult items
nocorall	Number correct for overall interview
ncreay	Number correct for easy items
ncrmod	Number correct for moderate items
nocrdif	Number correct for difficult items
xconfall	Mean confidence for overall interview
xcoesy	Mean confidence for easy items
xcomod	Mean confidence for moderate items
xcodiff	Mean confidence for difficult items
accy	Self-report; how accurate they thought they were
gdwtns	Self-report; how good a witness they thought they were
perfco	Self-report; how confident they were in their performance
evidcrt	Self-report; how happy they would be to give evidence in court

copecrt	Self-report; how well they thought they might cope with giving evidence in court
gdwitcrt	Self-report; how good a witness they thought they would be in court
hitsall	Hits for overall interview
hitsdiff	Hits for difficult items
flsalmall	False-alarms for overall interview
flsalmdif	False-alarms for difficult items
%accall	Percentage accuracy for overall interview
%accdif	Percentage accuracy for difficult items
acc5	Accuracy for overall interview at 5 minutes
acc6	Accuracy for overall interview at 6 months
conf5	Confidence for overall interview at 5 minutes
conf6	Confidence for overall interview at 6 months
caall5	Within-subjects confidence-accuracy correlation for overall interview at 5 minutes
caeasy5	Within-subjects confidence-accuracy correlation for easy items at 5 minutes
caeasy6	Within-subjects confidence-accuracy correlation for easy items at 6 months
camod5	Within-subjects confidence-accuracy correlation for moderate items at 5 minutes
camod6	Within-subjects confidence-accuracy correlation for moderate items at 6 months
cadiff5	Within-subjects confidence-accuracy correlation for difficult items at 5 minutes
cadiff6	Within-subjects confidence-accuracy correlation for overall interview at 6 months
discft5	Scale; measure of discomfort at retrieval at 5 minutes
discft6	Scale; measure of discomfort at retrieval at 6 months

switches	Mean number of switches on FAS (phonemic)
totwfas	Mean total word score on FAS (phonemic)
xclusfas	Mean cluster score on FAS (phonemic)
acceasy5	Accuracy for easy items at 5 minutes
acceasy6	Accuracy for easy items at 6 months
accmod5	Accuracy for moderate items at 5 minutes
accmod6	Accuracy for moderate items at 6 months
accdiff5	Accuracy for difficult items at 5 minutes
accdiff6	Accuracy for difficult items at 6 months
coneasy5	Confidence for easy items at 5 minutes
coneasy6	Confidence for easy items at 6 months
conmd5	Confidence for moderate items at 5 minutes
conmd6	Confidence for moderate items at 6 months
condif5	Confidence for difficult items at 5 minutes
condif6	Confidence for difficult items at 6 months
cocoall5	Confidence in correct answers for overall interview at 5 minutes
cocoall6	Confidence in correct answers for overall interview at 6 months
cocoesy5	Confidence in correct answers for easy items at 5 minutes
cocoesy6	Confidence in correct answers for easy items at 6 months
ccomod5	Confidence in correct answers for moderate items at 5 minutes
ccomod6	Confidence in correct answers for moderate items at 6 months
cocodif5	Confidence in correct answers for difficult items at 5 minutes
cocodif6	Confidence in correct answers for difficult items at 6 months
coinall5	Confidence in incorrect answers for overall interview at 5 minutes

coinall6	Confidence in incorrect answers for overall interview at 6 months
cinesy5	Confidence in incorrect answers for easy items at 5 minutes
cinesy6	Confidence in incorrect answers for easy items at 6 months
coinmd5	Confidence in incorrect answers for moderate items at 5 minutes
coinmd6	Confidence in incorrect answers for moderate items at 6 months
coindif5	Confidence in incorrect answers for difficult items at 5 minutes
coindif6	Confidence in incorrect answers for difficult items at 6 months
sracc5	Self-report; how accurate they thought they were at 5 minutes
sracc6	Self-report; how accurate they thought they were at 6 months
srconp5	Self-report; how confident they were in their performance at 5 minutes
srconp6	Self-report; how confident they were in their performance at 6 months
srgdwt5	Self-report; how good a witness they thought they were at 5 minutes
srgdwt6	Self-report; how good a witness they thought they were at 6 months
srcopct5	Self-report; how well they thought they might cope with giving evidence in court at 5 minutes
srcopct6	Self-report; how well they thought they might cope with giving evidence in court at 6 months
srevdct5	Self-report; how happy they would be to give evidence in court at 5 minutes
srevdct6	Self-report; how happy they would be to give evidence in court at 6 months
srgdwtct5	Self-report; how good a witness they thought they would be in court at 5 minutes

srgdwtct6	Self-report; how good a witness they thought they would be in court at 6 months

THE UNIVERSITY
of LIVERPOOL

DEPARTMENT OF PSYCHOLOGY RESEARCH ETHICS COMMITTEE

FORM 1: Ethical Approval for Research with Human Participants

All research activity conducted by any member of the Department of Psychology under the auspices of the Department of Psychology must comply with the British Psychological Society's guidelines regarding Ethical Principles for Conducting Research with Human Participants. Before conducting any such research, therefore, please read carefully and complete the following form with regard to your proposed research. N.B. IF YOU INTEND TO WORK IN SCHOOLS YOU MUST ALSO COMPLETE AND ATTACH FORM 2: ETHICAL APPROVAL FOR WORKING IN SCHOOLS.

a) **Taught students**: With regard to all research conducted by undergraduates, and on taught Master's courses, it is up to the individual supervisor/s to ensure that the research proposal complies with the guidelines. In cases of doubt, the proposal should be brought to the attention of the Chair of the Research Ethics Committee. ***When students submit practical research work, this form, signed by both student and supervisor, must be included.***

b) **Research students and staff**: With regard to all research conducted by research postgraduates and staff, ***the prospective researchers must return this form, together with an account of their proposed research, to the Chair of the Research Ethics Committee confirming that their proposed research conforms with the BPS guidelines.*** If the proposal is approved, the Chair will issue a formal letter to this effect. Normally this will be straightforward, however, in cases of doubt, or where there is good reason to think that the guidelines are infringed, the Chair of the Research Ethics Committee will consult other members to advise whether ethical approval should be granted, and, if necessary, what amendments are required to obtain approval.

Informed Consent

Normally the informed consent of all prospective participants in the research must be obtained. The investigator should, therefore, inform all participants of all aspects of the research or intervention that might reasonably be expected to influence the participants' willingness to participate. Sometimes, as in certain types of observational research, and in the consideration of some case record data, it may not be possible to obtain the individual consent of all participants. In such cases, wherever possible, a suitable subsample of the prospective participants should be consulted to ensure that, in principle, prospective participants would be likely to offer their consent. Observational research is only acceptable in situations where those observed would expect to be observed by strangers. Case record data can only be used with the consent of those authorized to release such records. If there is any predictable likelihood of physical or psychological harm or discomfort to participants as a result of the research procedures, real informed consent must always be obtained (see below under Protection of participants).

Deception

Intentional withholding of information from participants should be avoided whenever possible. Participants should never be deliberately misled without strong scientific justification. When it is considered that there are exceptional scientific reasons for misleading participants and/or

withholding information a suitable subsample of the prospective participants should be consulted to assess how these procedures may be received. It should be emphasised that, the withholding of information from, and the misleading of participants, are unacceptable if the participants are likely to object or show unease once debriefed.

Debriefing
In studies in which the participants are aware that they have taken part in an investigation the investigator should clearly and sincerely offer to a) answer participants' questions with regard to the nature of the research, and b) discuss with the participants their experience of the research to monitor any unforseen negative effects and misconceptions.

Withdrawal from the investigation
In studies in which the participants are aware that they are involved in an investigation, they must be told at the onset that they have a right to withdraw at any time regardless of whether payment or other inducement has been offered. Participants should also be informed that they may refuse to comply with any aspect of the research should they wish (for example, they may refuse to answer certain questions).

Confidentiality
Subject to the requirements of legislation, including the Data Protection Act, information obtained about a participant during an investigation is confidential unless the participant has agreed otherwise in advance.

Protection of participants
Investigators have a primary responsibility to protect participants from physical and mental harm during the investigation. Typically the risk of harm must be no greater than an average person might encounter in his or her normal life. Participants must be asked about any factors in the procedures that might present risks for them, such as a pre-existing medical condition, and must be advised of any special action they should take to avoid such risk. When the procedures are deemed to bear risks greater than those encountered in ordinary life, the investigator must obtain the approval of independent advisors (the Department of Psychology Research Ethics Committee in the first instance). The real informed consent of all participants must be obtained if there is any unusual risk of harm or discomfort from the procedures. Participants should also be informed of procedures for contacting the investigator if concerns arise despite these principles.

Confirmation of accordance with ethical principles
I/we confirm that the proposed research conforms with the guidelines as stated in this document with regard to ALL of the following: Informed Consent, Deception, Withdrawal from research, Debriefing, Confidentiality, and Protection of participants.

Signed (as applicable)

Student..... JACQUELINE WHEATCROFT .....

Supervisor/tutor/member of staff.... G.F. Wheatcroft .....

Chair of Research Ethics Committee... G.F. Wheatcroft .....

Date..... 20/3/02 .....